**Description of Facilities and Resources**

**Oak Ridge National Laboratory**
**and the**
**UT-ORNL Joint Institute for Computational Sciences**

## 1. Oak Ridge National Laboratory

**Computer Facilities.** The Oak Ridge National Laboratory (ORNL) hosts three petascale computing facilities: the Oak Ridge Leadership Computing Facility (OLCF), managed for DOE; the National Institute for Computational Sciences (NICS) computing facility operated for the National Science Foundation (NSF); and the National Climate-Computing Research Center (NCRC), formed as collaboration between ORNL and the National Oceanographic and Atmospheric Administration (NOAA) to explore a variety of research topics in climate sciences. Each of these facilities has a professional, experienced operational and engineering staff comprising groups in high-performance computing (HPC) operations, technology integration, user services, scientific computing, and application performance tools. The ORNL computer facility staff provides continuous operation of the centers and immediate problem resolution. On evenings and weekends, operators provide first-line problem resolution for users with additional user support and system administrators on-call for more difficult problems.

**Other Facilities.** The Oak Ridge Science and Technology Park at ORNL is the nation's first technology park on the campus of a national laboratory. The technology park is available for private sector companies that are collaborating with research scientists. Laboratory officials anticipate that the new park will be used to help create new companies from technologies developed at ORNL.

### *1.1 Primary Systems*

***Jaguar*** is a Cray XK6 system consisting of 18,688 AMD sixteen-core Opteron processors providing a peak performance of more than 3.3 petaflops (PF) and 600 terabytes (TB) of memory. A total of 384 service input/output (I/O) nodes provide access to the 10 PB "Spider" Lustre parallel file system at more than 240 gigabytes (GB/s). External login nodes (decoupled from the XK6 system) provide a powerful compilation and interactive environment using dual-socket, twelve-core AMD Opteron processors and 256 GB of memory. Jaguar also includes 960 NVIDIA Tesla M2090 Graphics Processing Units designed to accelerate calculations. Jaguar is the Department of Energy's most powerful open science computer system and is available to the international science community through the INCITE program, jointly managed by DOE's Leadership Computing Facilities at Argonne and Oak Ridge National Laboratories.


***Titan*** will be an upgrade to Jaguar in late 2012. Titan will add next generation NVIDIA GPUs to the nodes of Jaguar resulting in a system with a peak performance of between 10 and 20 PF. The Spider disk subsystem will be upgraded to provide up to 1 TB/s of disk bandwidth and up to 30 PB of storage.

***Gaea*** is a Cray XE6 system delivered in two stages. The first stage, delivered in the summer of 2010, consists of 2,576 socket G34 AMD 12-core Magny-Cours Opteron processors, providing 30,912 compute cores, 82.4 TB of double data rate 3 (DDR3) memory, and a peak performance of 260 teraflops (TF). The second stage consists of 4,896 socket G34 AMD 16-core Interlagos Opteron processors, providing 78,336 compute cores, 156.7 TB of DDR3 memory, and a peak performance of 721 TF.


After the stage two system enters production, the original stage one system will receive an architectural upgrade to the Interlagos processor. The resulting aggregate system will provide 1.106 PF of computing capability, and 248 TB of memory. The Gaea compute partitions are supported by a series of external login nodes and two separate file systems. The FS file system is based on more than 2,000 SAS drives and provides more than 1 PB (formatted) space for fast scratch to all compute partitions. The LTFS file system provides more than 2000 SATA drives and 4 PB formatted capacity as a staging and archive file system. Gaea is the NOAA climate community's most powerful computer system and is available to the climate research community through the Department of Commerce/NOAA.

***The ORNL Institutional Cluster (OIC)*** consists of two phases. The original OIC consists of a bladed architecture from Ciara Technologies called VXRACK. Each VXRACK contains two login nodes, three storage nodes, and 80 compute nodes. Each compute node has dual Intel 3.4 GHz Xeon EM64T processors, 4 GB of memory, and dual gigabit Ethernet interconnects. Each VXRACK and its associated login and storage nodes are called a block. There are a total of nine blocks of this type. Phase 2 blocks were acquired and brought online in 2008. They are SGI Altix machines. There are two types of blocks in this family.

- Thin Nodes (3 blocks). Each Altix contains 1 login node, 1 storage node, and 28 compute nodes within 14 chassis. Each node has eight cores and 16 GB of memory. The login and storage nodes are XE240 boxes from SGI. The compute nodes are XE310 boxes from SGI.
- Fat Nodes (2 blocks). Each Altix contains 1 login node, 1 storage node, and 20 compute nodes within 20 separate chassis. Each node has eight cores and 16 GB of memory. These XE240 nodes from SGI contain larger node-local scratch space and a much higher I/O to this scratch space because the space is a volume from four disks.

***Frost*** (SGI Altix ICE 8200) consists of three racks totaling 128 compute nodes, 5 service nodes (1 batch node and 4 login nodes), 2 rack leader nodes, and 1 administration node. Each compute node has two Intel quad-core Xeon X5560 at 2.8 GHz (Nehalem) processors, 24 GB of memory, a 1 Gb Ethernet connection, and two 4x DDR Infiniband connections. Each rack of compute nodes contains eight Infiniband switches (Mellanox InfiniScale III MT47396, 24 10-Gb/s Infiniband 4X ports) that are used as the primary interconnect between compute nodes and for connection to the Lustre file system. The center-wide Lustre file system is the main storage available to the compute nodes. The Frost cluster is available to ORNL staff and collaborators.

### *1.2 The University of Tennessee*

***Kraken*** is a Cray XT5 system consisting of 18,816 AMD six-core Opteron processors providing a peak performance of 1.17 PF and 147 TB of memory. It is connected to more than 3 PB of disk space for scratch space. At the current time, it is the eleventh fastest computer in the world, the fastest academic computer in the world, and the largest resource on the NSF XSEDE network.


### *1.3 Joint Institute for Computational Sciences*

The University of Tennessee (UT) and Oak Ridge National Laboratory (ORNL) established the Joint Institute for Computational Sciences (JICS) in 1991 to encourage and facilitate the use of high-performance computing in the state of Tennessee. When UT joined Battelle Memorial Institute in April 2000 to manage ORNL for the Department of Energy (DOE), the vision for JICS expanded to encompass becoming a world-class center for research, education, and training in computational science and engineering. JICS advances scientific discovery and state-of-the-art engineering by

- taking full advantage of the computers at the petascale and beyond housed at ORNL and in the Oak Ridge Leadership Computing Facility (OLCF) and
- enhancing knowledge of computational modeling and simulation through educating a new generation of scientists and engineers well versed in the application of computational modeling and simulation to solving the world's most challenging scientific and engineering problems.

**Joint Institute for Computational Sciences.**

JICS is staffed by joint faculty who hold dual appointments as faculty members in departments at UT and as staff members in ORNL research groups. The institute also employs professional research staff, postdoctoral fellows and students, and administrative staff.

The JICS facility represents a $10M investment by the state of Tennessee and features a state-of-the-art interactive distance learning center with seating for 66 people, conference rooms, informal and open meeting space, executive offices for distinguished scientists and directors, and incubator suites for students and visiting staff.

The JICS facility is a hub of computational and engineering interactions. Joint faculty, postdocs, students, and research staff share the building, which is designed specifically to provide intellectual and practical stimulation. The auditorium serves as the venue for invited lectures and seminars by representatives from academia, industry, and other laboratories, and

the open lobby doubles as casual meeting space and the site for informal presentations and poster sessions, including an annual 200+ student poster session.

In June 2004, JICS moved into a new 52,000 $ft^2$ building next door to the OLCF. The OLCF, which is located on the ORNL campus, is among the nation's most modern facilities for scientific computing. The OLCF includes 40,000 square feet divided equally into two rooms designed specifically for high-end computing systems.

Within JICS, there are three other centers that are the result of three large National Science Foundation (NSF) awards:

The National Institute for Computational Sciences (NICS) at the University of Tennessee is the product of a $65M NSF Track 2B award. The mission of NICS is to enable the scientific discoveries of researchers nationwide by providing leading-edge computational resources and education, outreach, and training for underrepresented groups. Kraken, the fastest, most powerful supercomputer for academic use, is the flagship NICS computing resource.

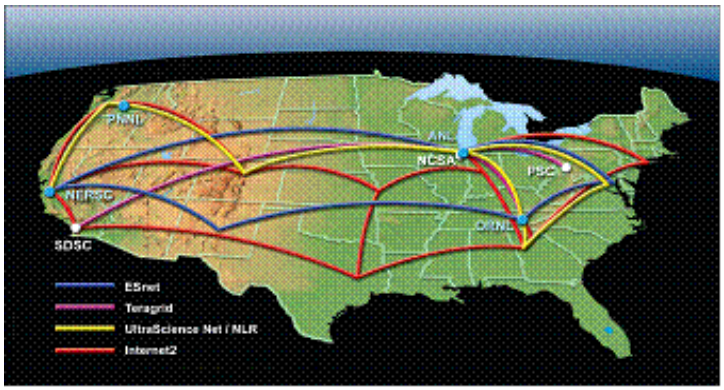
The UT Center for Remote Data Analysis and Visualization (RDAV) is sponsored by NSF through a 4-year, $10 million TeraGrid XD award. The centerpiece hardware resource at RDAV is Nautilus, a new SGI UltraViolet shared-memory machine featuring 1,024 cores and 4 terabytes of memory within a single system image. A wide range of software tools is available for TeraGrid users to perform data analysis, visualization, and scientific workflow automation on Nautilus. The machine is located at ORNL and is administered by NICS staff.

The Keeneland Project is a 5-year, $12 million Track 2 grant awarded by NSF for the deployment of an experimental high-performance system. The Georgia Institute of Technology and its project partners, UT-Knoxville and ORNL, have initially acquired and deployed a small, experimental, high-performance computing system consisting of an HP system with NVIDIA Tesla accelerators attached. The machine is located at ORNL and is administered by NICS staff.

## 2. Infrastructure

**Physical and Cyber Security.** ORNL has a comprehensive physical security strategy including fenced perimeters, patrolled facilities, and authorization checks for physical access. An integrated cyber security plan encompasses all aspects of computing. Cyber security plans are risk-based. Separate systems of differing security requirements allow the appropriate level of protection for each system, while not hindering the science needs of the projects.

**Network Connectivity.** The ORNL campus is connected to every major research network at rates of between 10 GB/s and 100 GB/s. Connectivity to these networks is provided via optical networking equipment owned and operated by UT-Battelle that runs over leased fiber-optic cable. This equipment has the capability of simultaneously carrying either 192 10-GB/s circuits or 96 40-GB/s circuits and connects the OLCF to major networking hubs in




**ORNL network connectivity to university, national laboratory, and industry partners.**

Atlanta and Chicago. Currently, 16 of the 10 GB circuits are committed to various purposes, allowing for virtually unlimited expansion of the networking capability. The connections into ORNL provide access to research and education networks such as ESnet, TeraGrid, Internet2, and Cheetah at 10 GB/s; Science Data Net at 20 GB/s; and National LambdaRail at 40 GB/s. ORNL operates the Cheetah research network for NSF. To meet the increasingly demanding needs of data transfers between major facilities, ORNL is participating in the Advanced Networking Initiative that provides a native 100 GB optical network in a loop which includes ORNL, Argonne National Laboratory, Lawrence Berkeley National Laboratory, and other facilities in the northeast.

**Students in the Research Alliance in Math and Science program experience the EVEREST power wall.**

The local-area network is a common physical infrastructure that supports separate logical networks, each with varying levels of security and performance. Each of these networks is protected from the outside world and from each other with access control lists and network intrusion detection. Line rate connectivity is provided between the networks and to the outside world via redundant paths and switching fabrics. A tiered security structure is designed into the network to mitigate many attacks and to contain others.

**Visualization and Collaboration.** ORNL has state-of-the-art visualization facilities that can be used on site or accessed remotely.

ORNL's **E**xploratory **V**isualization **E**nvironment for **RE**search in **S**cience and **T**echnology (EVEREST) is a 30-ft wide by 8-ft high power wall for data exploration and analysis. The facility has a 600 $ft^2$ projection area and a 1000 $ft^2$ viewing area known as the EVEREST lab, a venue that serves both as a visualization center and a place for scientists to meet, hold discussions, and present their work. The ORNL visualization team has developed a suite of middleware software tools that offers an intuitive interface with which to operate the power wall and manage multimedia content. Twenty-seven projections are seamlessly edge-matched for an aggregate resolution of 11,520 by 3,072 pixels. This projection environment is driven by an 18-node cluster named Everest. Each node in the Everest cluster contains four dual-core AMD Opteron processors, 4GB of memory, dual NVIDIA GeForce 8800GTX graphics cards, and an Infiniband network. A dedicated Lustre file system provides high bandwidth data delivery to the EVEREST power wall. ORNL also provides Lens, a 77-"fat node" cluster dedicated to data analysis and visualization. 45 nodes of Lens contain 16 AMD cores, 128 GB of memory and an Infiniband network. The remaining 32 nodes of Lens contain 16 AMD cores, 64 GB of memory, Infiniband network, and two graphics cards, an NVIDIA 8800 GTX and a 4GB NVIDIA Tesla C1060. The Lens cluster is a resource of the OLCF and performs a variety of visualization-related functions, including computation, analysis, and rendering, including support for remote visualization for off-site customers. The Lens cluster has been demonstrated with a variety of commercial off-the-shelf software and open-source visualization tools including VisIt, Paraview, CEI Ensight, and AVS-Express. The Everest cluster rendering environment utilizes Chromium and Distributed Multi-Head X (DMX) for tiled, parallel rendering. The Lens cluster cross mounts the Center-wide Lustre file system to allow "zero copy" access to simulation data from other OLCF computational resources.

**OLCF tape archive.**

**High Performance Storage and Archival Systems.** To meet the needs of ORNL's diverse computational platforms, a shared parallel file system capable of meeting the performance and scalability requirements of these platforms has been successfully deployed. This shared file system, based on Lustre, Data Direct Networks (DDN), and InfiniBand technologies, is known as Spider and provides centralized access to petascale datasets from all major on-site computational platforms. Delivering more than 240 GB/s of aggregate performance, scalability to more than 26,000 file system clients, and more than 10-petabyte (PB) storage capacity, Spider is the world's largest scale Lustre file system. Spider consists of 48 DDN 9900 storage arrays managing 13,440 1-TB SATA drives; 192 Dell dual-socket, quad-core I/O servers providing more than 14 TF in performance; and more than 3 TB of system memory. Metadata are stored on 2 LSI Engino 7900s (XBB2) and are served by three Dell quad-socket, quad-core systems. ORNL systems are interconnected to Spider via an InfiniBand system area network which consists of four 288-port Cisco 7024D IB switches and more than 3 miles of optical cables. Archival data are stored on the center's High Performance Storage System (HPSS), developed and operated by ORNL. HPSS is capable of archiving hundreds of petabytes of data and can be accessed by all major leadership computing platforms. Incoming data are written to disk and later migrated to tape for long term archiving. This hierarchical infrastructure provides high-performance data transfers while leveraging cost effective tape technologies. Robotic tape libraries provide tape storage. The center has three SL8500 tape libraries holding up to 10,000 cartridges each and is deploying a fourth SL8500 in 2011. The libraries house a total of 24 T10K-A tape drives (500 GB cartridges, uncompressed) and 32 T-10K-B tape drives (1 terabyte cartridges, uncompressed). Each drive has a bandwidth of 120 MB/s. ORNL's HPSS disk storage is provided by DDN storage arrays with nearly a petabyte of capacity and over 12 GB/s of bandwidth. This infrastructure has allowed the archival system to scale to meet increasingly demanding capacity and bandwidth requirements with more than 21 PB of data stored as of November 2011.