CADES: Accelerating scientific discovery through integrated compute and data solutions

CADES At-A-Glance

- Fully integrated infrastructure offering compute and data for ORNL researchers
- Offers computing, data analytics, data storage and networking to facilitate scientist's data workflows
- User facilities, staff members, and research teams can process, manage, and analyze large amounts of data by leveraging CADES resources
- Currently being used by researchers across ORNL to address big data challenges

Modern science depends on computing. In fact, many recent scientific discoveries would not be possible without the aid of computers and specialized instruments. But as scientific computing has advanced, our ability to manage and extract insights from the resulting massive amounts of data has struggled to keep pace. An expanded high-performance computing ecosystem is called for to address scientists' data needs.

Introducing CADES, the Compute and Data Environment for Science at the US Department of Energy's (DOE's) Oak Ridge National Laboratory (ORNL). Launched internally by the Computing and Computational Sciences Directorate, CADES is a fully integrated infrastructure offering compute and data services for researchers lab-wide. Be it computing, data analytics, data storage, or networking, CADES provides a one-stop shop for everything researchers need to maximize scientific output.

By facilitating scientists' data workflows, CADES serves as a foundational "big data center" that aspires to address present and future DOE data challenges. With this platform, user facilities, staff members, and research teams can process, manage, and analyze large amounts of data by leveraging CADES high-performance computing resources, scalable storage, data analysis, and visualization tools.

The CADES infrastructure supports multiple data security levels with separate open research and secure network enclaves. The open research enclave includes small-scale high-performance computing (HPC) and cloud compute systems and interconnects with DOE's Energy Sciences Network (ESnet) at 100 Gigabits per second (Gbps). Over five petabytes (PBs) of scratch storage and project workspace is available through Spider, ORNL's Lustre-based parallel file system. The secure enclave for proprietary information supports similar capabilities while meeting required certification and accreditation standards.

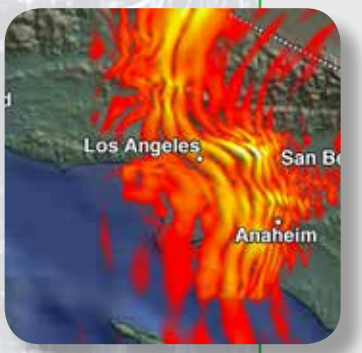
Instead of a one-size fits all program, CADES offers users a range of options, so research teams can customize their computing and data environment to best fit the needs of their science program. In the day-to-day operation of CADES, computing and computational sciences staff members provide as-needed project support based on their expertise in areas such as HPC, cloud computing, networking, and visualization.

ORNL researchers are currently using CADES to facilitate lab projects from a variety of disciplines.

At the Spallation Neutron Source (SNS), which produces neutrons to study the atomic structure of materials, CADES users are shortening the time it takes to analyze their data by leveraging the Accelerating Data Acquisition, Reduction, and Analysis (ADARA) system. ADARA integrates experimental data and computational analysis through a live-streaming infrastructure that provides on-the-spot results to researchers at an SNS workstation.

Other use cases have resulted in customized, large-scale resources, including:

- a web portal infrastructure that provides data on atmospheric carbon compounds recorded from a network of ground-based spectrometers for the Total Carbon Column Observing Network, a partnership between ORNL's Carbon Dioxide Information Analysis Center and the California Institute of Technology.
- the deployment of large memory systems for image analysis at ORNL's Center for Nanophase Materials Sciences.
- extreme data analysis for computationally demanding cosmology projects.
- the coupling of CADES resources to the Titan supercomputer to support advanced climate modeling with an automated workflow, post-processing of large datasets, and visualization through web-based analytic tools.

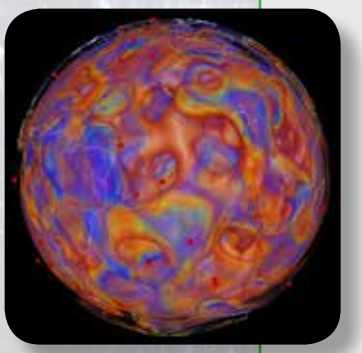


COMPUTE

In addition to the Cray XK7 Titan supercomputer, CADES provides users with a number of HPC resources.

Metis is a hybrid GPU–CPU Cray XK7 system that gives users access to a Titan-like environment that exists in an open research protection zone. Kos is a flexible open research resource with over 2,300 cores and 5.4 terabytes (TBs) of memory. Similarly, Niyol has 1,100 cores and 2.4 TBs of memory for proprietary research users needing the moderate protection zone. Additionally, CADES offers the ORNL Institutional Cluster, or OIC, comprised of over 10,500 compute cores for modeling and simulations, along with access to their Geographic Information System, or GIS cloud.

DATA STORAGE

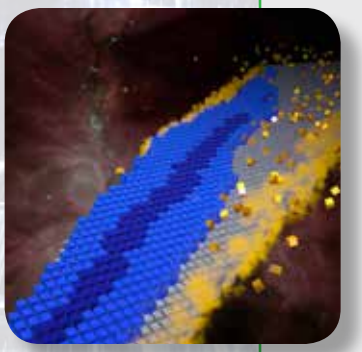


To ensure data has a place to stay, users have several options for storage. For open research, users can access Dasein, a large-scale, high-performance storage system with over 4 petabytes (PBs) of storage, or Azhi, a mid-scale, high-performance storage environment with 1 PB of storage and moderate protection.

DATA ANALYTICS

The CADES infrastructure delivers several avenues for processing data via analytics or visualization tools.

For data analytics under moderate protection zones CADES offers the Apollo and Helios systems. Helios, a Cray Hadoop platform, offers 504 cores with 5.4 TBs of memory and 576 TBs for local storage. Additionally, the Cray Urika Apollo appliance has 64 Threadstorm4 processor nodes with 2 TBs of RAM, supporting up to 8,192 hardware threads, and 125 TBs of local Lustre storage—scalable to 512 TBs.



NETWORKING

No matter where data comes from or where it goes, the CADES data transfer node infrastructure provides high-performance data-motion through ESnet, the world's fastest open science network, connecting users to all 17 U.S. national labs, as well as over 140 research institutions worldwide.

- Large-scale InfiniBand networking backplane with over 24 Terabit/sec of bisectional bandwidth
- Large-scale Ethernet networking backplane with over 6 Terabites/sec of bisectional bandwidth

For more information about CADES **visit www.cades.ornl.gov**, or contact Arjun Shankar at **shankarm@ornl.gov**.