High End Computing (HEC) Infrastructure and Applications (I&A)

NITRD Agencies: NSF, NIH, DOE/SC, NASA, NIST, OSD, NOAA, EPA, DOE/NNSA

HEC I&A members coordinate Federal activities in advanced computer systems, applications software, and related infrastructure, which have become core necessities for cutting-edge discovery across all scientific and engineering fields. HEC capabilities provide researchers with an unprecedented ability to study the most complex entities and processes in biology, chemistry, environmental sciences, materials science, nanoscale science and technology, physics, and many other fields. In HEC I&A, Federal agencies work collaboratively to extend the state of the art in computing systems, science and engineering applications, and data management. Amid increasing global competition, ongoing improvements in U.S. high-end computing capabilities are needed to keep the Nation at the forefront of 21st century science and engineering advances and to increase our understanding in such critical areas as air and water quality, biological systems at all scales, energy sources and conservation, environmental and human-engineered systems, and weather and climate.

President's 2006 Request

Strategic Priorities Underlying This Request

- Sustaining U.S. scientific and engineering leadership in certain areas that are supported by Federal agencies' science and national security missions will require ongoing investment in Federal HEC facilities and advanced computational applications over the next five years. When extrapolating the recent pace of advances in the use of available computing power to the end of the decade, petascale systems that are more than 100 times more powerful than today's highest-end systems may well be necessary to help meet mission needs and maintain the U.S. at the forefront of computational science innovation. The following critical issues documented in the "Federal Plan for High-End Computing" – developed by the High-End Computing Revitalization Task Force (HECRTF) and presented to Congress by the Administration in May 2004 – are being addressed in HEC agencies' leadership computing and acquisition coordination activities in HEC I&A:
  - *Availability and accessibility of Federal HEC resources for leading-edge scientific researchers.* During 2005 and 2006, new high-end capacity systems will be made available to the scientific research community (e.g., NSF, DOE/SC) that will expand researchers' access to HEC resources.
  - *Next-generation HEC platforms for the largest-scale scientific problems.* A new generation of capability platforms will be provided to the scientific and engineering communities (e.g., NASA, NSF, DOE/SC, DOE/NNSA). An example of a capability platform is NASA's new Columbia system, whose overall high performance enables scientists to execute in a day calculations that would previously have taken months (e.g., the simulations validating the most likely cause of the wing damage that led to the Columbia shuttle disaster).
  - *Coordinated, streamlined mechanisms for benchmarking and procurement of Federal HEC systems.* The agencies are developing mechanisms for opening computing cycles on leadership-class systems to peer-reviewed projects from the broad academic and industrial R&D communities, and are working collaboratively on new performance benchmarks to more accurately predict computing system performance on diverse scientific problems and the total cost of ownership.
- The modeling and simulation of complex scientific systems and engineering problems enabled by high-end computing are now an essential form of research and development. Recent successes, such as NOAA's ability for the first time to accurately predict 72 hours in advance the unusual track of hurricane Charlie in September 2004, saving lives and property, illustrate the great positive impact of computational science on the Nation. Such examples of national impact, many cited in the HECRTF report, motivate HEC I&A priorities.

Highlights of Request

Acquisition of prototype leadership-class machines

- **NSF:** Pittsburgh Supercomputing Center's new 2,000-processor, 10-teraflop Red Storm machine will be available for research by science discipline users requiring orders of magnitude increases in capability
- **DOE/SC (ORNL):** Operation of two 20-teraflop peak capability leadership-class architectures to enable scientific discovery

- **NASA (ARC):** New 10,420-processor, massively parallel processing (MPP) Columbia system will be partly available for research by all science discipline users requiring orders of magnitude increases in capability over today's levels

Applications
- **DOE/SC:** Scientific Discovery Through Advanced Computing (SciDAC) program's achievement in establishing multidisciplinary, multi-institutional teams of computer and disciplinary scientists to develop high-quality software applications in the physical and biological sciences will be extended through re-competition of modeling and simulation applications for scientific discovery
- **DOE/SC:** Competition to select a small number of university-based SciDAC institutes that can become centers of excellence in high-end computational science in areas critical to DOE missions and software centers for high performance computing
- **DOE/SC:** Successful 2004 Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program (providing competitive access to 10 percent of National Energy Research Scientific Computing Center [NERSC] system) will be re-competed in 2006

Planning and Coordination Supporting Request
Infrastructure
- **DOE/SC, NOAA, NSF, OSD:** Connectivity and technology delivery to universities
- **DARPA, NASA, NIST, NSA, NSF:** Standards
- **NASA, OSD:** Acquisition coordination
- **DOE/NNSA, DOE/SC, NASA, NSA, NSF, OSD:** Capability community resources

Applications
- **DOE/SC, NASA, NIH, NSF:** Multiscale biological modeling
- **DOE/SC, EPA, NASA, NOAA, NSF, OSD:** Climate and weather applications
- **NASA, NOAA, NSF, OSD:** Weather research and forecasting
- **EPA, NOAA:** Air quality modeling

2005 and 2006 Activities by Agency

NSF: NSF Cyberinfrastructure Framework – supports supercomputing, data resources, and networking needs of the broad scientific community; Extensible Terascale Facility (ETF), with addition of Red Storm at Pittsburgh Supercomputing Center in mid-2005 (this acquisition complements the 25 teraflops already available in the ETF); ETF moves from construction phase to operational phase
NIH: National Centers for Biomedical Computing; Multiscale Modeling Initiative for Bioscience
DOE/SC: New capacity Linux cluster at NERSC; INCITE program; scientific applications pilots; multiscale mathematics, SciDAC re-competition; leadership-class computing
NASA: Ames Research Center – 52-teraflop Columbia system for aerospace and Earth science; funding for 2006 reflects strategic focus on large-scale system and reduction in investments in Grand Challenge Applications
NIST: IMPI; immersive visualization infrastructure; modeling of nanostructure of cement and concrete; fundamental mathematical tools (e.g., basic linear algebra subroutines [BLAS])
OSD (HPCMPO): Additional capacity for the High Performance Computing Centers (four major shared resource centers – ARL, ASC, ERDC, NAVO; 16 distributed centers); software applications support (Common HPC Software Support Initiative [CHSSI], HPC Software Applications Institutes, Programming Environment and Training, Software Protection Center)
NOAA: Facilities – Geophysical Fluid Dynamics Laboratory, Forecast Systems Laboratory, National Center for Environmental Protection; acquire and optimize the use of high-performance scalable systems for research
EPA: Air quality applications – optimization for scalable systems; GEOSS – enhanced environmental codes; metabonomics initiative – shared access to data and results; facilities – Grid deployment
DOE/NNSA: Capability systems – Purple, BlueGene/L, Red Storm; capacity systems – Linux HPC clusters; disruptive technologies – BlueGene/L as a scalable, low-power, low-cost-per-peak-TF computing system