The Federal Networking and Information Technology Research and Development Program

2006 NASA/IEEE Conference on
Mass Storage Systems and Technologies

May 17, 2006

Simon Szykman, Ph.D.

Director

**National Coordination Office for
Networking and Information Technology
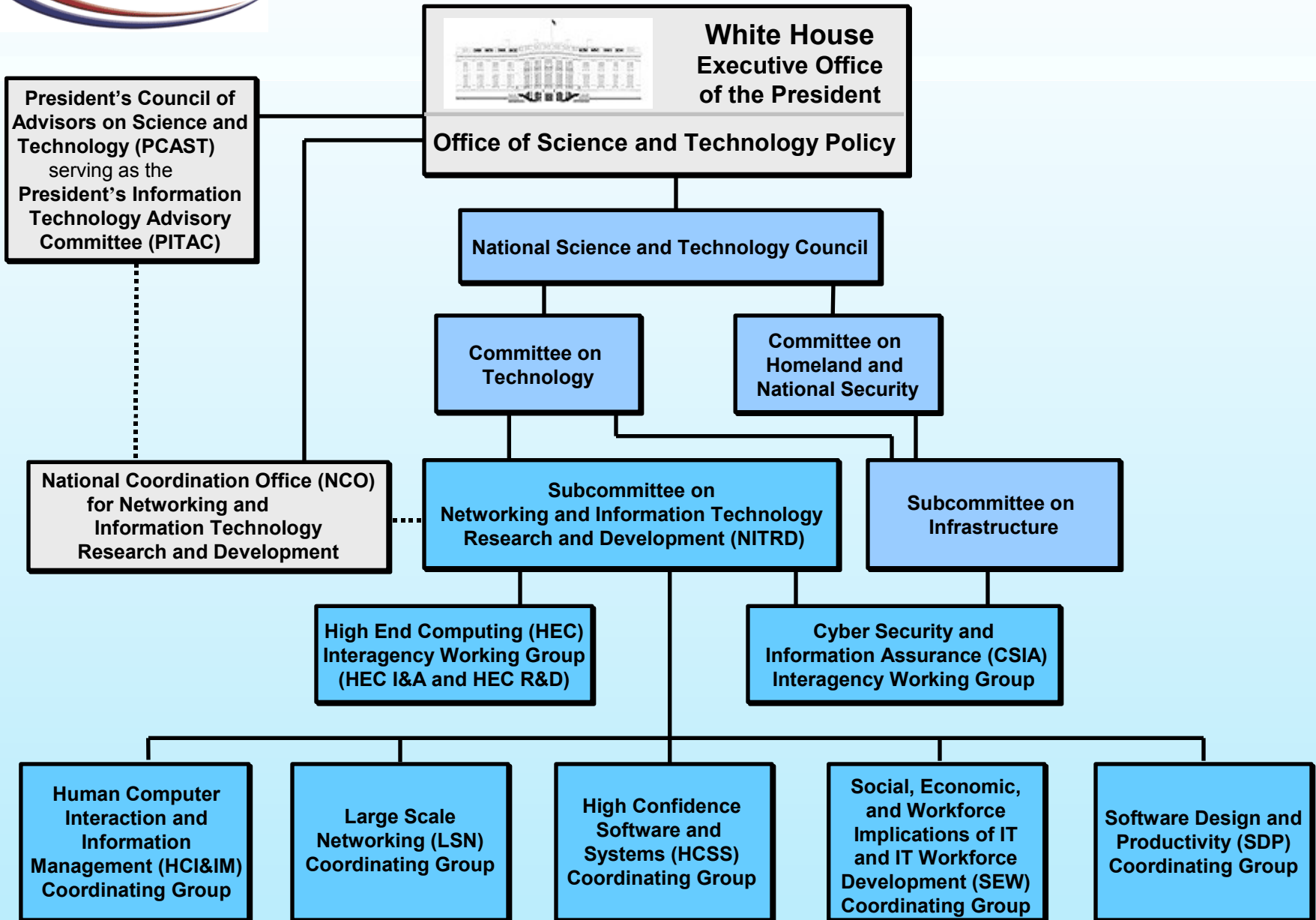Research and Development**

Overview of the NITRD Program

- **Authorization of the Networking and Information Technology Research and Development (NITRD) Program**
  - High-Performance Computing Act of 1991
  - Next Generation Internet Research Act of 1998
- **NITRD Subcommittee, National Science and Technology Council (NSTC)**
  - Representatives from 13 program agencies + OMB + OSTP + NCO/NITRD
  - Has two Interagency Working Groups (IWGs) and five Coordinating Groups (CGs)
- **Budget of $3.1 billion proposed for FY 2007**

NITRD Program Coordination Groups

White House
Executive Office
of the President

Office of Science and Technology Policy

President's Council of Advisors on Science and Technology (PCAST) serving as the **President's Information Technology Advisory Committee (PITAC)**

National Science and Technology Council

Committee on Technology

Committee on Homeland and National Security

National Coordination Office (NCO) for Networking and Information Technology Research and Development

Subcommittee on Networking and Information Technology Research and Development (NITRD)

Subcommittee on Infrastructure

High End Computing (HEC) Interagency Working Group (HEC I&A and HEC R&D)

Cyber Security and Information Assurance (CSIA) Interagency Working Group

Human Computer Interaction and Information Management (HCI&IM) Coordinating Group

Large Scale Networking (LSN) Coordinating Group

High Confidence Software and Systems (HCSS) Coordinating Group

Social, Economic, and Workforce Implications of IT and IT Workforce Development (SEW) Coordinating Group

Software Design and Productivity (SDP) Coordinating Group

3

NITRD NCO Objectives

- **Support NITRD-related policy making in the White House Office of Science and Technology Policy (OSTP)**

- **Serve as the Federal focal point for interagency technical planning, budget planning, and coordination for the Federal NITRD Program**

- **Serve as a source of timely, high-quality, technically accurate, in-depth information on accomplishments, new directions, and critical challenges for the NITRD Program**

Agency NITRD Budgets by PCA

FY 2007 Budget Requests (dollars in millions)

Agency		High End Computing Infrastructure & Applications (HEC I&A)	High End Computing Research & Development (HEC R&D)	Cyber Security & Information Assurance (CSIA)	Human-Computer Interaction & Information Management (HCI &IM)	Large Scale Networking (LSN)	High Confidence Software & Systems (HCSS)	Social, Economic, & Workforce Implications of IT (SEW)	Software Design & Productivity (SDP)	Total
NSF	2006 Estimate	220.3	62.7	57.6	207.4	82.2	41.3	91.1	47.9	810.3
	2007 Request	**272.4**	**64.1**	**67.6**	**220.9**	**84.0**	**51.3**	**92.9**	**50.7**	**903.7**
OSD & DoD Service research orgs.		214.6	9.8	0.6	138.5	141.8	31.2	0.2	6.9	543.7
		186.0	**8.7**	**0.7**	**135.6**	**130.7**	**29.1**	**0.3**	**6.8**	**497.8**
NIH		198.5			188.7	74.9	8.4	12.3	17.9	500.6
		194.7			**183.2**	**74.6**	**8.3**	**12.2**	**17.7**	**490.7**
DARPA			94.1	78.7	174.2	21.3				368.3
			117.7	**81.6**	**233.2**	**33.2**				**465.7**
DOE/SC		104.4	109.1			38.9		3.5		255.8
		135.3	**160.4**			**45.0**		**4.0**		**344.7**
NSA			89.2	14.1		1.0	36.2			140.5
			62.4	**13.3**		**2.3**	**39.9**			**117.9**
NASA		60.3		1.3	2.0	5.7	7.0		1.8	78.1
		63.9		**1.3**	**2.0**	**6.0**	**7.0**		**1.8**	**82.0**
AHRQ					40.1	21.6				61.7
					37.3	**20.0**				**57.3**
NIST		2.3	1.2	9.1	7.8	4.3	9.6		4.6	38.9
		2.3	**1.2**	**11.1**	**9.8**	**4.3**	**9.6**		**4.6**	**42.9**
DOE/NNSA		10.0	15.9			1.6		4.6	3.3	35.4
		9.5	**23.4**			**1.6**		**4.6**	**2.8**	**41.9**
NOAA		11.4	1.9		0.2	0.7			1.6	15.8
		16.4	**1.9**		**0.5**	**2.9**			**1.6**	**23.3**
EPA		3.3			3.0					6.3
		3.3			**3.0**					**6.3**
TOTAL (2006 Estimate)		825.0	383.9	161.3	761.9	393.9	133.6	111.6	84.0	2,855
TOTAL (2007 Request)		**883.8**	**439.9**	**175.5**	**825.4**	**404.5**	**145.2**	**114.0**	**85.9**	**3,074**

5

American Competitiveness Initiative (ACI)

- **Calls for a doubling over 10 years of the investment in three Federal agencies — NSF, DOE/SC, and NIST — that support basic research programs in the physical sciences and engineering**

- **All three agencies are NITRD Program members**

- **2007 budget increases for ACI agencies exceed the percentage increase in the overall proposed NITRD Program budget**
  - NSF: ⇨ 12%
  - DOE/SC: ⇨ 35%
  - NIST: ⇨ 10%
  - Collective increase for ACI agencies is $186 million (17% above 2006 estimates)
  - ACI agency budgets accounts for over 85% of the overall NITRD Program budget increase for 2007

MSST in the NITRD Program

- **Mass Storage Systems and Technologies are relevant to multiple NITRD Program Components Areas**
  - Human-Computer Interaction and Information Management (HCI&IM)
  - High End Computing Research and Development (HEC R&D)
  - High End Computing Applications and Infrastructure (HEC I&A)

- **HCI&IM**
    - HCI&IM R&D aims to increase the benefit of computer technologies to humans, particularly the science and engineering R&D community.
    - HCI&IM R&D invests in technologies for mapping human knowledge into computing systems, communications networks, and information systems and back to human beings, for human analysis, understanding, and use.
    - R&D areas include: cognitive systems, data analysis in fields such as human health and the environment, information integration, multimodal and automated language translation, robotics, and user interaction technologies.

- **Agency for Healthcare Research and Quality (AHRQ)**
- **Defense Advanced Research Projects Agency (DARPA)**
- **DoD Service research organizations**
- **Environmental Protection Agency (EPA)**
- **National Aeronautics and Space Administration (NASA)**
- **National Institutes of Health (NIH)**
- **National Institute of Standards and Technology (NIST)**
- **National Oceanic and Atmospheric Administration (NOAA)**
- **National Science Foundation (NSF)**
- **Other participating agencies**
  - General Services Administration (GSA)
  - National Archives and Records Administration (NARA)
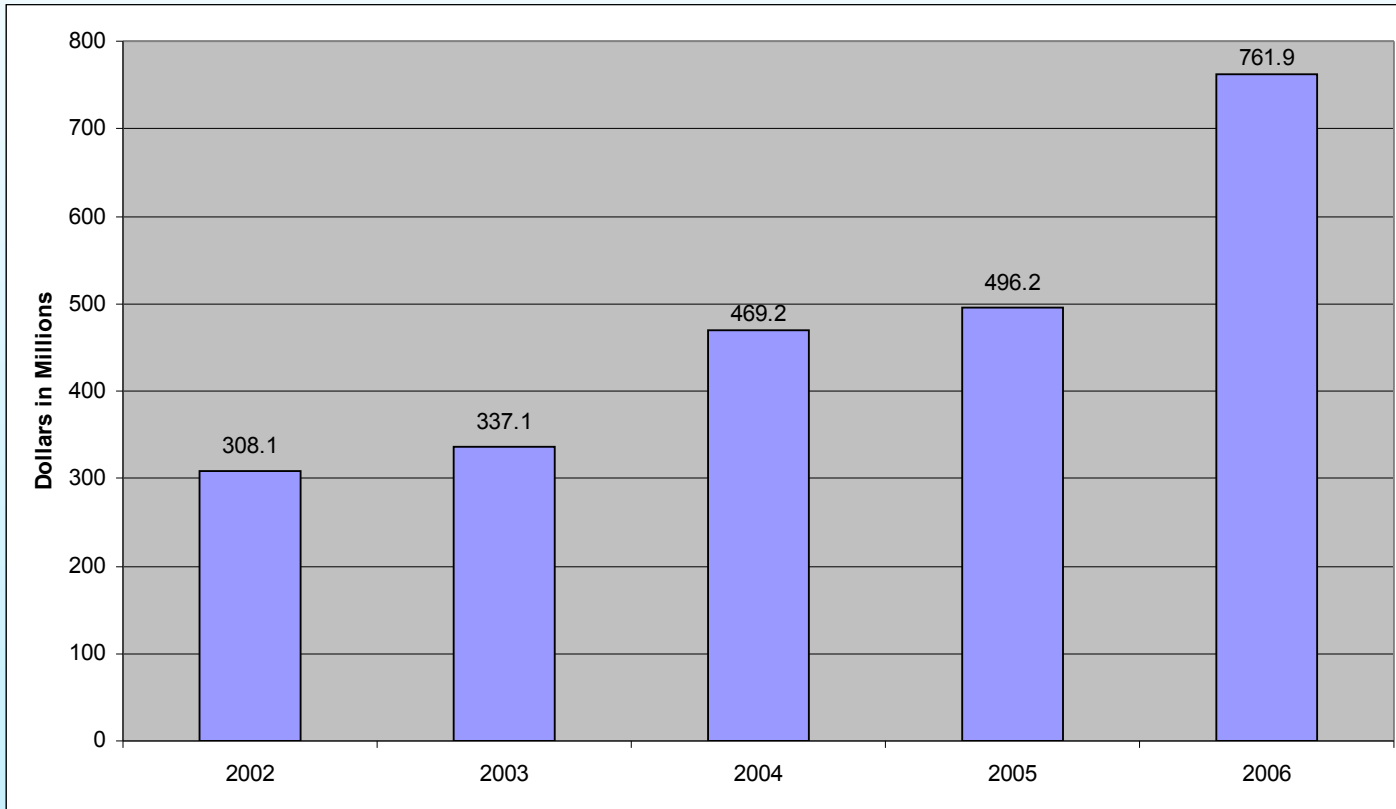
FY 2007 HCI&IM Budget Request: $825 Million

- **Information accessibility, integration, and management:**
  - Next-generation methods, tools, and technologies to make it possible to access, integrate, analyze, and efficiently manage massive stores of widely distributed, heterogeneous information, with the long-term goal of developing techniques that can be generalized across domains. Needs include:
    - Federal information management architecture testbeds: To evaluate issues in petascale collections of information governed by differing requirements (e.g., national security vs. personal privacy)
    - Long-term preservation: Maintenance of and access to long-lived science and engineering data collections and Federal records
- **Multimodal devices and interfaces:**
  - Human-computer interaction capabilities enabling rapid, easy access to and communication and understanding of heterogeneous information
- **Systems that know what they are doing:**
  - Cognitive systems that are able to adjust to change, and repair themselves

- **Data-intensive discovery and design environments:**
  - Interdisciplinary environments leveraging large-scale data exploration, analysis, and understanding, including concurrent visualization, algorithms and tools (NASA)
- **Remote Sensing Information Gateway:**
  - Global Earth Observation Systems of Systems (GEOSS) demonstration project to share and integrate Earth observational data (EPA, with NASA, NIH, NOAA)
- **Data security and data analysis methods:**
  - Research in analysis of digital images and videos; research in methods for computational analysis of data collected in the observational sciences (NSF)
- **Multimodal language recognition and translation and Global Autonomous Language Exploitation (GALE):**
  - Improved performance and evaluation of human language technologies (DARPA, NSA, NSF, NIST, DTO, NARA and others)

- **National workshop on information integration R&D:**
  - Issues for coordinated research include interoperability, privacy, security, and standards (NSF, AHRQ, DoD, EPA, NARA, with NIST, GSA, and others)

- **Drug information and standards:**
  - Build system to obtain drug information with standardized definitions and in standardized formats (AHRQ, NIH, NIST, FDA, HHS, and others)

- **Earth System Modeling Framework:**
  - Information interoperability and reuse in Earth science applications (NASA, DOE/SC, NOAA, NSF, OSD and DoD Service research organizations, and others)

- **Eco-Informatics:**
  - Workshop and plans for possible second joint solicitation (NSF, NASA, EPA, and others)

- **Health informatics:**
  - Planning for collaboration to include workshop and joint program activities (NSF, NIH)

- **Support for empirical research and evaluations of cluster disk storage architectures to address IM challenges of massive volumes of data.**

  - IM architectures where mass storage research is fused with experiments, evaluations, and tests that contribute to understanding across the systems stack of applications, file systems, and storage (NARA, NSF)

  - Exemplar: Transcontinental Persistent Archives Prototype – A research test bed used to address the Nation's challenge of safeguarding, preserving, and providing access to electronic records (NARA, NSF)

- **Scalable grid-associated mass storage to understand I/O and alternatives for optimization of file systems and mass storage (NARA, NSF)**

Applications Fueling Research in Mass Storage

- **Management of very large datasets, use of metadata, and development of decision support tools for knowledge discovery and data display (NOAA)**

- **Tools and approaches to explore potential linkages between air quality and human health (EPA, NOAA)**

- **Integration of search and retrieval techniques across environmental and health libraries (EPA, NOAA)**

Definition of the HEC PCAs

- **HEC R&D**
  - HEC R&D agencies conduct and coordinate hardware and software R&D to enable the effective use of high-end systems to meet Federal agency mission needs, to address many of society's most challenging problems, and to strengthen the Nation's leadership in science, engineering, and technology.
  - Research areas of interest include hardware (e.g., microarchitecture, memory subsystems, interconnect, packaging, I/O, and storage), software (e.g., operating systems, languages and compilers, development environments, algorithms), and systems (e.g., system architecture, programming models).

- **HEC I&A**
  - HEC I&A agencies coordinate Federal activities to provide advanced computing systems, applications software, data management, and HEC R&D infrastructure to meet agency mission needs and to keep the United States at the forefront of 21st century science, engineering, and technology.
  - HEC capabilities enable modeling and simulation of complex processes in biology, chemistry, climate and weather, environmental sciences, materials science, nanoscale science and technology, physics, and other areas to address Federal agency mission needs.
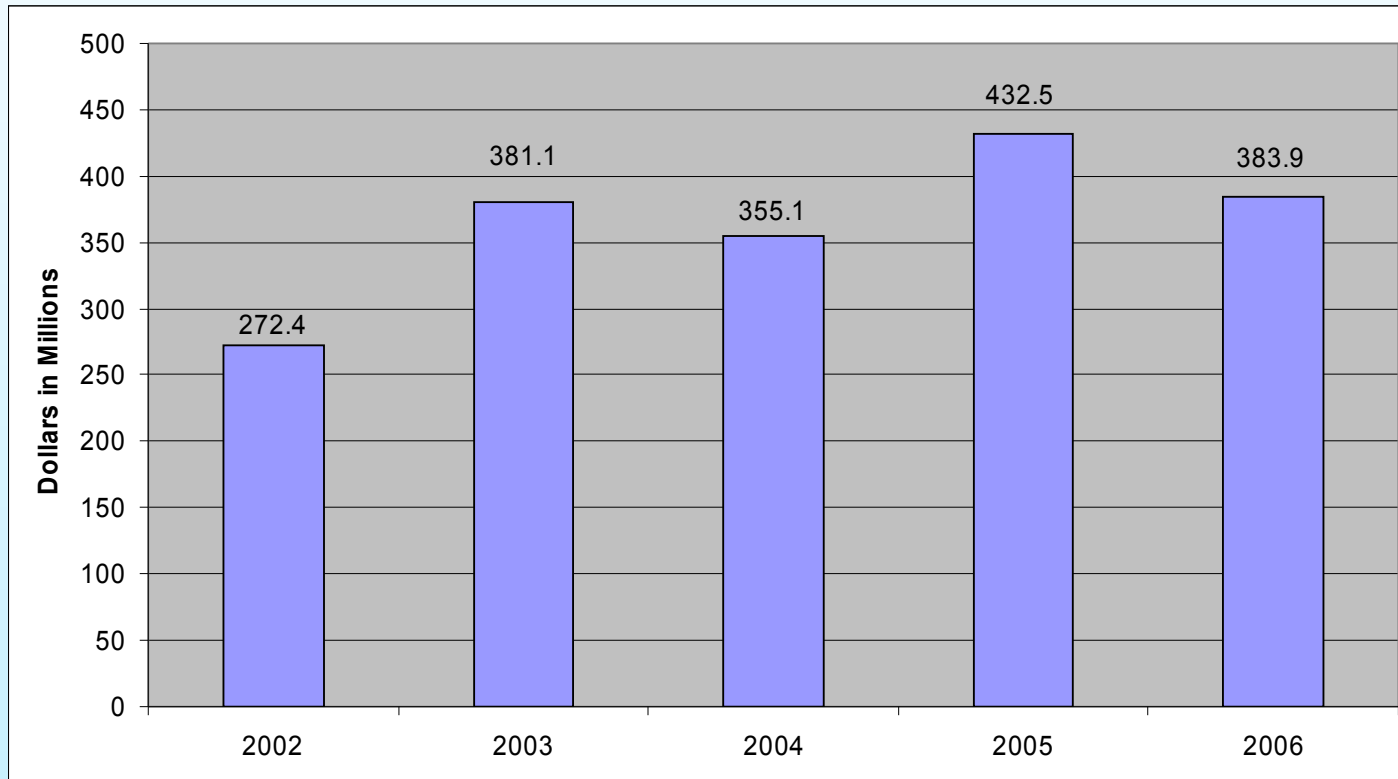
HEC Agencies in the NITRD Program

- **Defense Advanced Research Projects Agency (DARPA)**
- **Department of Energy National Nuclear Security Administration (DOE/NNSA)**
- **Department of Energy Office of Science (DOE/SC)**
- **Environmental Protection Agency (EPA)**
- **National Aeronautics and Space Administration (NASA)**
- **National Institute of Standards and Technology (NIST)**
- **National Institutes of Health (NIH)**
- **National Oceanic and Atmospheric Administration (NOAA)**
- **National Science Foundation (NSF)**
- **National Security Agency (NSA)**
- **Office of the Secretary of Defense (OSD) and DoD Service research organizations**
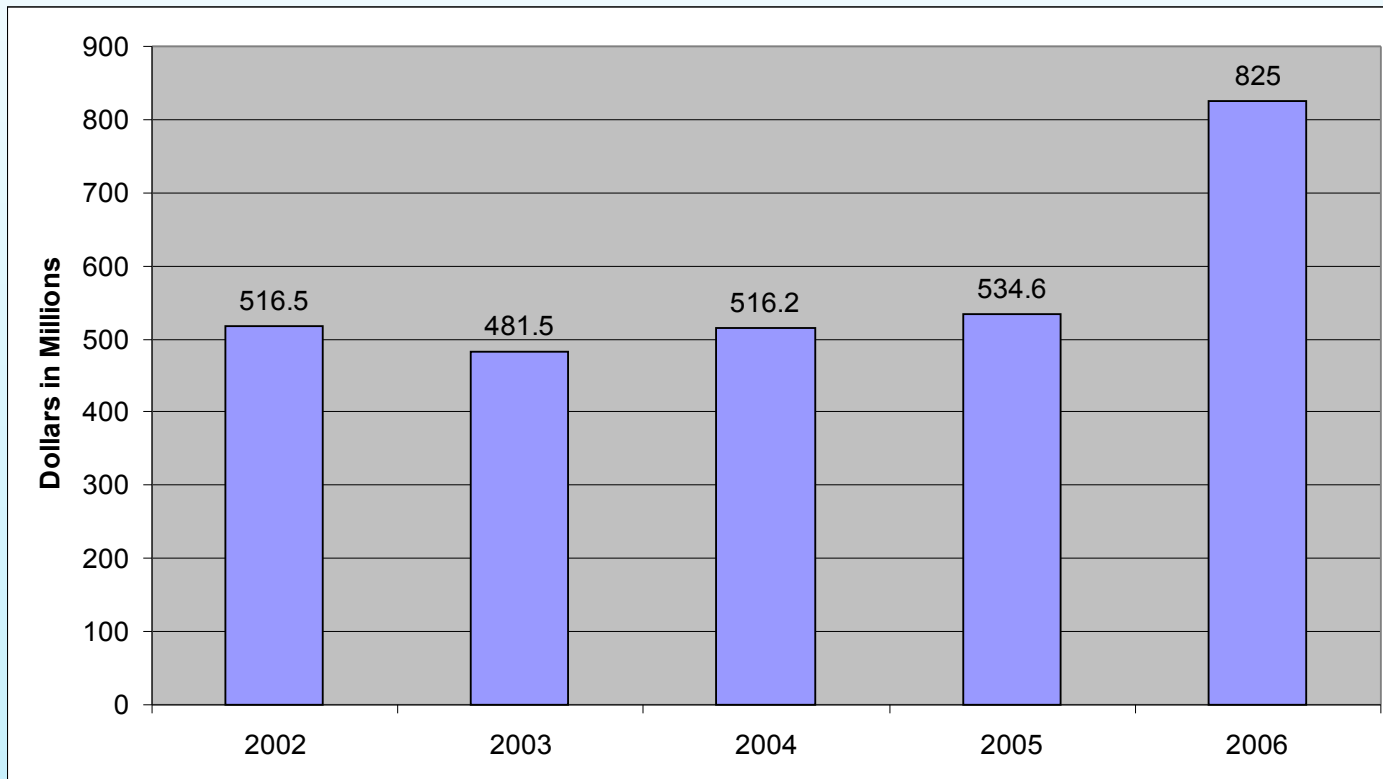 - DoD HPC Modernization Program Office (HPCMPO)

FY 2007 HEC R&D Budget Request: $440 Million
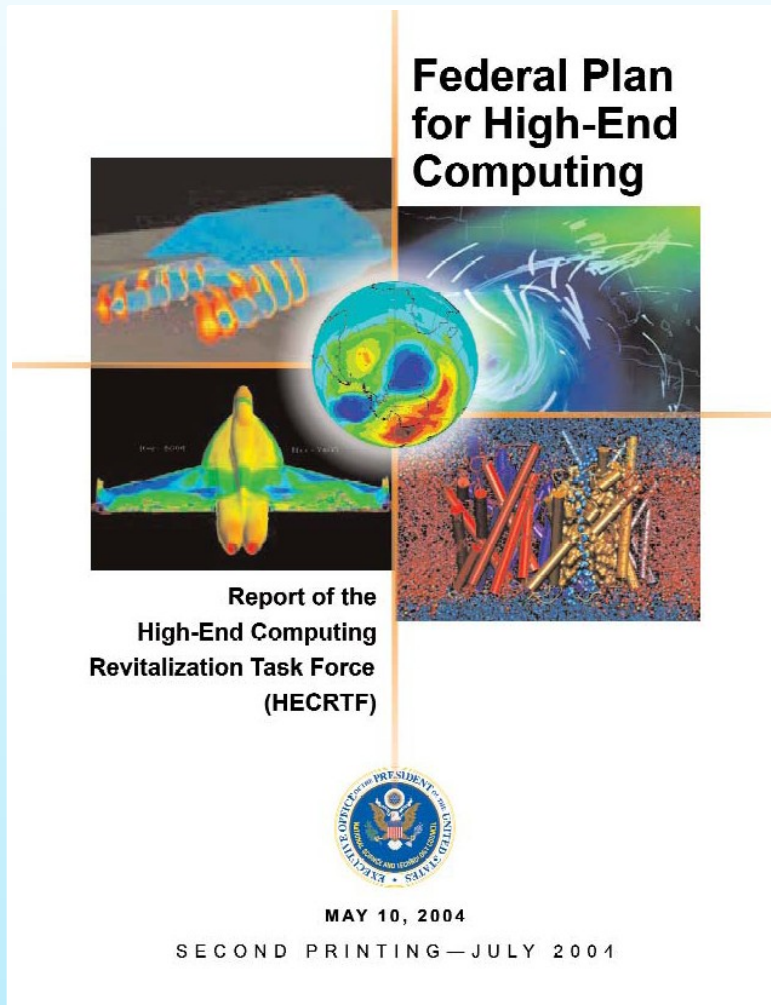
FY 2007 HEC I&A Budget Request: $884 Million

Federal Plan for High-End Computing

Report of the High-End Computing Revitalization Task Force (HECRTF)

MAY 10, 2004
SECOND PRINTING—JULY 2001

- R&D in high-end computing
  - I/O and storage is an element of the hardware R&D Roadmap
- Resources
  - Production systems
  - Leadership systems
  - Accessibility
- Procurement practices

HEC Priorities and Activities

- **HEC R&D**
  - DARPA High Productivity Computing Systems (HPCS) Program
    - HPCS vendor systems, productivity research, joint program reviews (DOE/SC, DOE/NNSA, NSA, DoD/HPCMPO, NSF)
  - Benchmarking and performance measurement
  - I/O and file systems
  - Petascale system software
  - HEC University Research Activity (HEC-URA)
    - NSF, DOE/SC, DARPA, NSA funded research in operating/ runtime systems, programming models, performance modeling and optimization, data management
- **HEC I&A: leadership class computing**
  - Petascale systems and applications
- **Procurement and acquisition**
  - Several agencies sharing/borrowing/adapting acquisition practices (NSF, DoD/HPCMPO, DOE/SC, and DOE/NNSA)

- **Leadership class facilities**
  - Upgrades and new platforms at DOE/SC (ORNL, ANL, LBNL), NASA (ARC, GSFC) and NSF
- **Applications**
  - Parallel & distributed algorithms
  - Fundamental mathematical tools
  - Expand the predictive regimes of weapons and engineering codes
  - DOE/SC Scientific Discovery Through Advanced Computing (SciDAC) program, to extend SciDAC's multidisciplinary, multi-institutional teams to petascale open science applications
- **Maintain a production-level computational infrastructure**
- **HPCS Program, Phase III development and demonstration**
  - Petascale application development
  - Productivity measurement and evaluation process
  - High productivity language system environments

Highlights of HEC Joint or Multiagency Planning and Coordination

- **Benchmarking (DARPA, DoD/HPCMPO, NSF, NASA, NOAA, DDR&E, EPA)**
- **Acquisition Process (DoD/HPCMPO, NSF, DOE/SC, DOE/NNSA, NASA, NOAA, DDR&E)**
- **System Allocation (DOE/SC, NSF, NASA)**
- **Cooperative Platform Development (DOE/SC, NSA, DOE/NNSA)**
- **Earth System Modeling Framework (ESMF) / Weather Research and Forecast (WRF) (NASA, DoD, DOE/SC, NOAA, NSF, EPA)**
- **HEC-URA (NSF, DOE/SC, DOE/NNSA, NSA, DARPA)**
- **HPCS Program, Phase III (DARPA, DOE/SC, NASA, DOE/NNSA, NSA, NSF, DoD/HPCMPO)**
- **File Systems and I/O (DOE/NNSA, DoD, NSF, DOE/SC, NSA, NASA, NRO)**
- **Council on Competitiveness HPC Initiative (DARPA, DOE/SC, DOE/NNSA)**
- **Project Reviews (DOE/NNSA, DOE/SC, NSF)**
- **Computational Neuroscience (NIH, NSF)**
- **Multi-Scale Modeling (NIH, NSF)**

- **Comparison of rate of change of technologies**

Technology		Growth rate
Transistors per integrated circuit	k_{moore}	0.46
LINPACK on Top 10 supercomputer	k_{top500}	0.58
Capacity of hard drives	k_{IO_cap}	0.62
Cost per GB of storage	k_{IO_cost}	-0.84
Performance of hard drive in IO's per second	$k_{IO\_perf\_io}$	0.20
Performance of hard drive in bandwidth	$k_{IO\_perf\_B}$	0.26

- **As technologies advance, where will bottlenecks occur in HEC systems and applications?**

Source: Subramaniyan R., Studham S., Grobelny E., "Optimization of Checkpointing-related I/O for High-Performance Parallel and Distributed Computing," Submitted to the 2006 International Conference on Parallel and Distributed Processing Techniques and Applications.

HEC File Systems and I/O Research Guidance Workshop

- **HPCS Workshop on I/O Requirements (DARPA)**
  - Held July 15, 2005
  - Agencies discussed HEC I/O requirements
    - DOE/NNSA, NSF, DOE/SC, NSA, DoD/HPCMPO, NRO
- **HEC File Systems and I/O Research Guidance Workshop**
  - Held August 16-17, 2005
  - Goals of workshop:
    - Catalog existing government-funded and other relevant research
    - List top research areas (both short and long term) that need to be addressed
    - Determine where gaps and overlaps exist
    - Recommend the most pressing future short and long term research areas and needs and other actions necessary to ensure well coordinated government-funded research

- Both evolutionary and revolutionary research into metadata issues

- Measurement and understanding of end-to-end I/O performance

- QoS throughout the hardware and software I/O stack

- Aspects of security such as usability, long term key management, and distributed authentication

- Next-generation I/O architectures

- Remote Direct Memory Access (RDMA), Object Based Secure Disk (OBSD) extensions, Network File System Version 4 (NFSv4) extensions, and parallel Network File System (pNFS)

- Management and reliability, availability, and serviceability (RAS) at scale

High End Computing University Research Activity (HEC-URA)

- Led by NSF, with co-funding from DOE, DARPA, NSA
- FY 2006 call for proposals in areas that include:

  - **File systems research**
  - **Quality of Service**
  - **Security**
  - **I/O middleware**
  - **Archives and backups as extensions to file systems**
  - **Novel storage devices for the I/O stack**
  - **I/O architectures**
  - **Management, reliability and availability at scale**

  - **Future file-systems-related protocols**
  - **Hardware and software tools for design and simulation of I/O, file, and storage systems**
  - **Efficient benchmarking, tracing, performance measurement, and tuning tools of I/O, file, and storage systems**

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=13645&org=CCF

Comments or Questions?

- **More detailed information is available in The FY 2007 Supplement to the President's Budget for the NITRD Program**

- **Visit http://www.nitrd.gov/**
- **Send e-mail to nco@nitrd.gov**
- **Call us at (703) 292-4873**