NASA Advanced Supercomputing Facility Expansion

Industry Day

21 February 2017

Background

The NASA Advanced Supercomputing (NAS) Division enables advances in high-end computing technologies and in modeling and simulation methods to tackle some of the toughest science and engineering challenges facing NASA today.

The name "NAS" has long been associated with leadership and innovation throughout the high-end computing (HEC) community. We play a significant role in shaping HEC standards and paradigms, and provide leadership in the areas of large-scale InfiniBand fabrics, Lustre open-source filesystems, and hyperwall technologies.

We provide an integrated high-end computing environment to accelerate NASA missions and make revolutionary advances in science. Pleiades, a petaflop-scale supercomputer, is used by scientists throughout the U.S. to support NASA missions, and is ranked among the most powerful systems in the world.

One of our key focus areas is in modeling and simulation to support NASA's real-world engineering applications and make fundamental advances in modeling and simulation methods.

Goals

Provide effective, production-level HEC resources and services to enable significant, timely impacts across NASA's mission areas of Earth and space science, aeronautics research, and human and robotic space exploration.

Advance the state-of-the-art in HEC technologies and techniques to meet NASA's continuously growing requirements for advanced computational modeling, simulation, and analysis.

Design and conduct high-fidelity numerical simulation studies for NASA's aerospace engineering programs, supporting both mission-critical and system design decisions.

Pursue fundamental advances in numerical methodologies and algorithms, physical model enhancements, and application code development for large-scale simulations of interest to NASA.

Integrate and operate IT systems to enable successful missions for specific NASA science and engineering projects.

Where We Fit In NASA

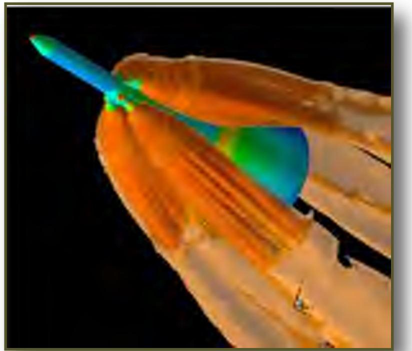
The NAS Division is part of the Exploration Technology Directorate at the NASA Ames Research Center. The Directorate’s mission is to create innovative and reliable technologies for NASA missions.

NAS operates NASA’s High-End Computing Capability Project, which is funded through the agency’s High-End Computing Program and the Space Environments Testing Management Office.

HECC Requirements

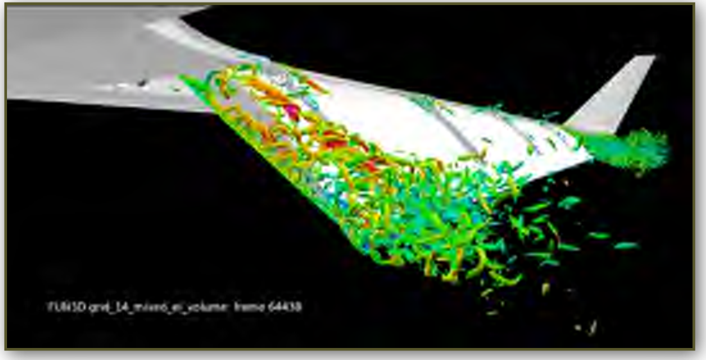
NASA's HEC Requirements: Capacity

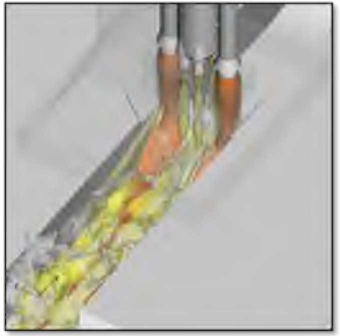
HEOMD (engineering-related work) require HEC resources that can handle large numbers of relatively-low CPU-count jobs with quick turnaround times.



Over 1500 simulations utilized ~ 2 million processor hours to study launch abort systems on the next generation crew transport vehicle

The formation of vortex filaments and their roll-up into a single, prominent vortex at each tip on a Gulfstream aircraft

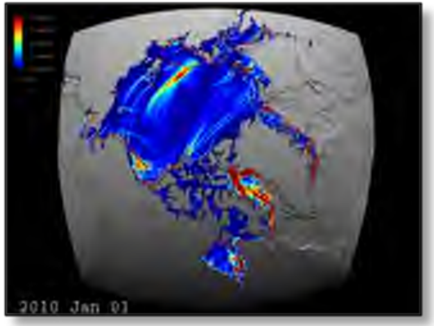




Over 4 million hours were used over a 4 month project to evaluate future designed of the next generation launch complex at the Kennedy Space Center

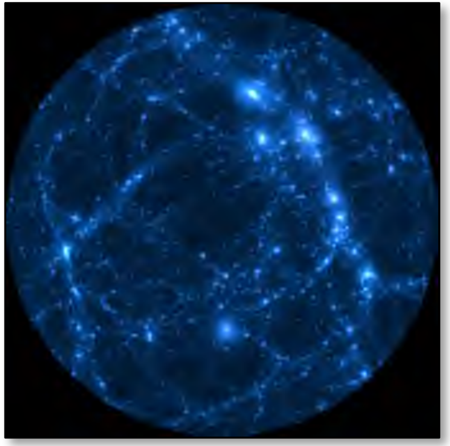
NASA's HEC Requirements: Capability

ARMD and SMD (aeronautics and science related work) require HEC resources that can handle high fidelity relatively-large CPU-count jobs with minimal time-to-solution. Capability enables work that wasn't possible on previous architectures.



NASA is looking at the oceans, running 100's of jobs on Pleiades using up to 10,000 processors. Looking at the role of the oceans in the global carbon cycle is enabled by access to large processing and storage assets

For the first time, the Figure-of-Merit has been predicted within experimental error for the V22 Osprey and Black Hawk helicopter rotors in hover, over a wide range of flow conditions



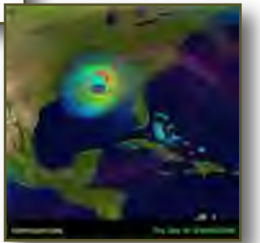
To complete the Bolshoi simulation, which traces how the largest galaxies and galaxy structures in the universe were formed billions of years ago, astrophysicists ran their code for 18 straight days, consuming millions of hours of computer time, and generating massive amounts of data

NASA's HEC Requirements: Time Critical

NASA also has need for HEC resources that can handle time-sensitive mission-critical applications on demand (maintain readiness)

HECC enables the enormous planetary transit searches to be completed in less than a day, as opposed to more than a month on the Kepler SOC systems, with significantly improved accuracy and effectiveness of the software pipeline

REENTRY

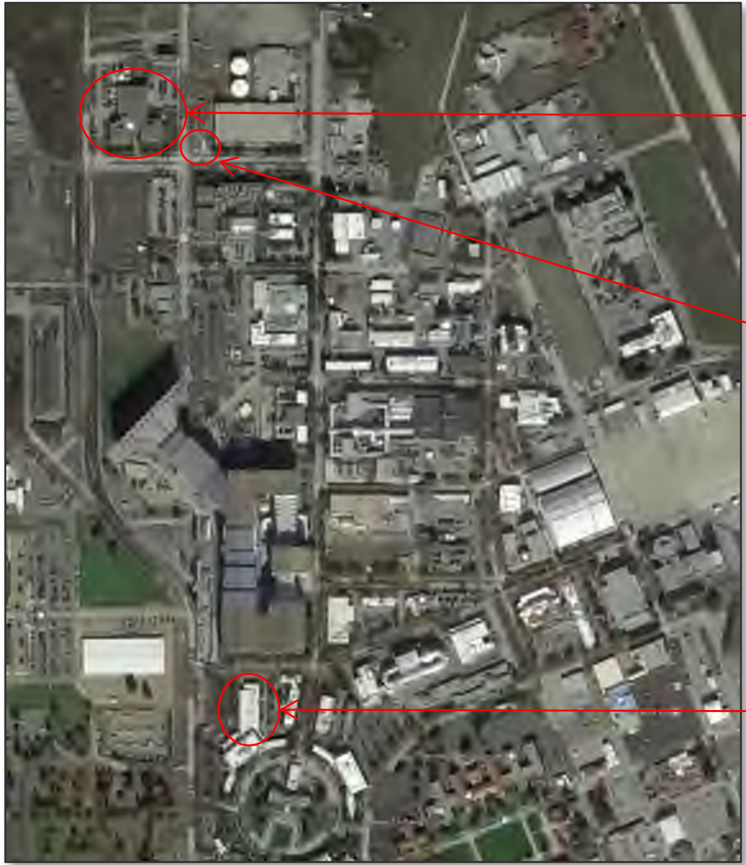


STORM PREDICTION

UAVSAR produces polarimetric (PolSAR) and interferometric (repeat-pass InSAR) data that highlight different features and show changes in the Earth over time

HECC Current Assets

Facilities

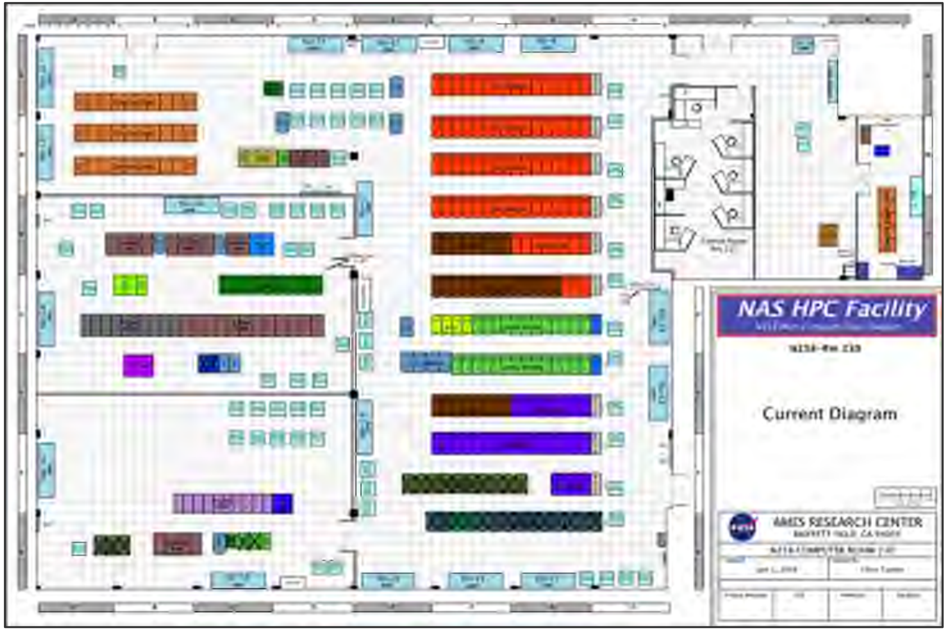


N258

R&D088

N233A

HECC Traditional Computer Floors

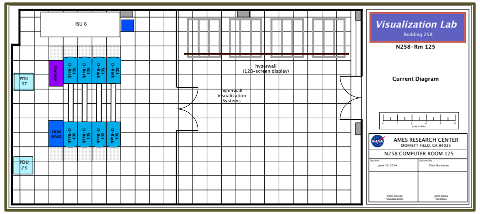


230
16,800

131
2,080

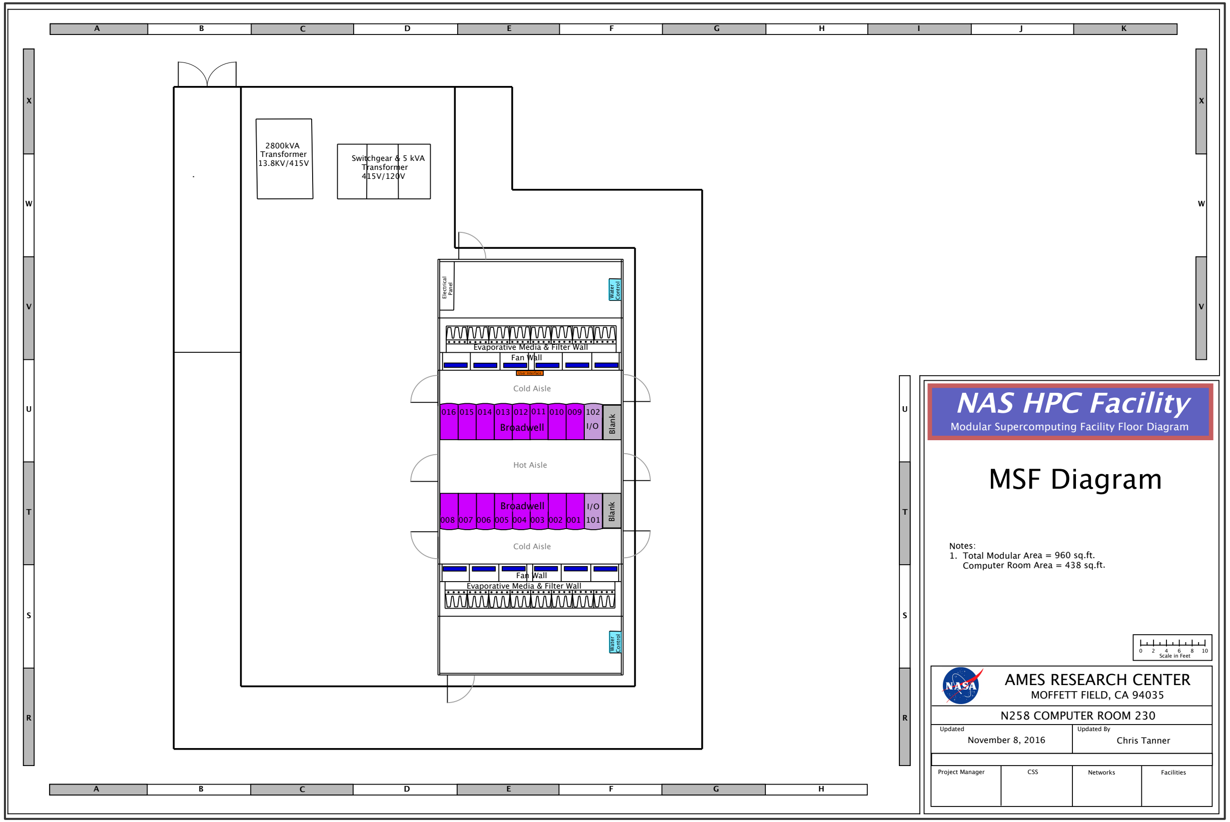
190
6,900

189
2,700



125
1,275

HECC Modular Computer Floors



R&D 088
16,800

Pleiades Specifics

161 SGI Racks (7.58 PF; 936 TB; 32,308 SBUs/hr)*

158 SGI Altix ICE X Racks (7.21 PF; 931 TB; 32,134 SBUs/hr)

- 26 racks of ICE-X with Intel Xeon processor E5-2670 (Sandy Bridge):
  623 TF; 59.9 TB; 3,407 SBUs/hr
- 75 racks of ICE-X with Intel Xeon processor E5-2680v2 (Ivy Bridge):
  2.419 PF; 345.6 TB; 13,608 SBUs/hr
- 29 racks of ICE-X with Intel Xeon processor E5-2680v3 (Haswell):
  2.004 PF; 267.3 TB; 6,974 SBUs/hr
- 28 racks of ICE-X with Intel Xeon processor E5-2680v4 (Broadwell):
  2.167 PF; 129.0 TB; 8,145 SBUs/hr

3 SGI Coyote Racks (371 TF; 5 TB; 175 SBUs) (note, accelerators are not counted in SBU numbers)

- 2 racks of Intel Xeon processor E5-2670 (Sandy Bridge) and Nvidia K40 graphic processors: 296 TF; 4 TB; 117 SBUs
- 1 rack of Intel Xeon processor E5-2670 (Sandy Bridge) and Intel Xeon Phi 5110P accelerator processor: 75 TF; 1 TB; 58 SBUs

Cores

- 22,944 Intel Xeon processors (246,048 cores)
- 64 Nvidia GPUs (184,320 cores)
- 64 Intel Xeon Phi processors (3,804 cores)

Nodes

- 11,376 nodes (dual-socket blades)
- 64 nodes (dual-socket + GPU)
- 32 nodes (dual-socket + dual-Phi)
- 14 Login nodes

Networks

- Internode: Dual-plane partial 11D hypercube (FDR)
- Gigabit Ethernet Management Network

** 1 SBU equals 1 hour of a Pleiades Westmere 12-core node.*

Electra Specifics

16 SGI Racks (1.24 PF; 147 TB; 4,654 SBUs/hr)

- 16 racks of ICE-X with Intel Xeon processor E5-2680v4 (Broadwell): 1.24 PF; 147 TB; 4,654 SBUs/hr

Cores

- 2,304 Intel Xeon processors (32,256 cores)

Nodes

- 1,152 nodes (dual-socket blades)

Networks

- Internode: Dual-plane fully-populated 7D hypercube (FDR)
- Gigabit Ethernet Management Network
- Metro-X IB extenders for shared storage access

Merope Specifics

56 SGI Altix ICE X ½ Racks (252 TF; 86 TB; 1,792 SBUs)

- 56 ½-racks of 8400EX with Intel Xeon processor E5670 (Westmere): 252 TF; 86 TB; 1,792 SBUs

3,584 Intel Xeon processors (21,504 cores)

- 3,584 six-core Westmere
- 2.93 GHz processors (21,504 cores)

Endeavour Specifics

32 TF constellation-class supercluster

2 SGI Ultra Violet 2 nodes with Intel Xeon E5-4650L 2.6 GHz processors

- One 512-core node with 2 TB globally addressable RAM (Endeavour1)
- One 1,024-core node with 4 TB globally addressable RAM (Endeavour2)

Interconnect

- Intranode: NUMALink-6 (enable large SSI)
- Dual-Plane QDR InfiniBand connectivity into Pleiades infrastructure
  - 1 connection from each node into IB0 for TCP traffic (pbs, login, …)
  - IB1 is for I/O traffic to the Lustre file systems. Endeavour1 has 3 connections and Endeavour2 has 4 connections.
- 10 Gb Ethernet can be used for WAN traffic

Advanced Visualization: hyperwall and CV

Supercomputing-scale visualization system to handle massive size of simulation results and increasing complexity of data analysis

- 7 8x16 LCD tiled panel display (23 feet x 10 feet)
- 7 245 million pixels
- 7 Debuted as #1 resolution system in the world
- 7 In-depth data analysis and software

Two primary modes

- 7 Single large high definition image
- 7 Sets of related images (e.g. parameter study)

High-bandwidth to HEC resources

- 7 Concurrent Visualization: Runtime data streaming allows visualization of every simulation time step - ultimate insight into simulation code without increase in traditional disk I/O
- 7 Traditional Post-Processing: Direct read/write access to Pleiades filesystems eliminates nee for copying large datasets

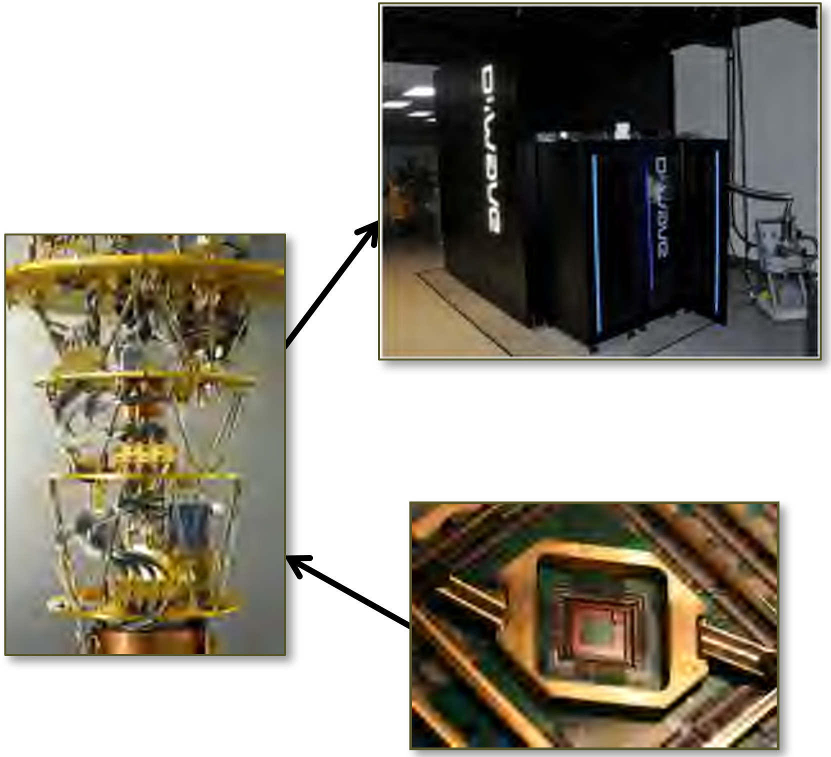
GPU-based computational acceleration R&D for appropriate NASA codes

Quantum Computing: D-Wave Two™ System

Collaboration between NASA / Google / USRA

D-Wave 2X Installed at NAS

- 7 Washington processor – 1,097 qubits (quantum bits – niobium superconducting loops encoding 2 magnetic states)
- 7 Physical characteristic
  - » 10 kg of metal in vacuum at 15 mK
  - » Magnetic shielding to 1 nanoTesla (50,000x less than Earth's magnetic field)
  - » Uses 12 kW electrical power
- 7 Focused on solving discrete optimization problems via quantum annealing



Storage and Archive

Lustre File Systems (39.6 PB in 7 file systems)

- DDN
  - 14 DDN RAID Systems, 9.9 PB total, 3 facility-wide file systems
- NetApp
  - 62 RAID Systems, 29.7 PB total, 4 facility-wide file systems

NFS File Systems

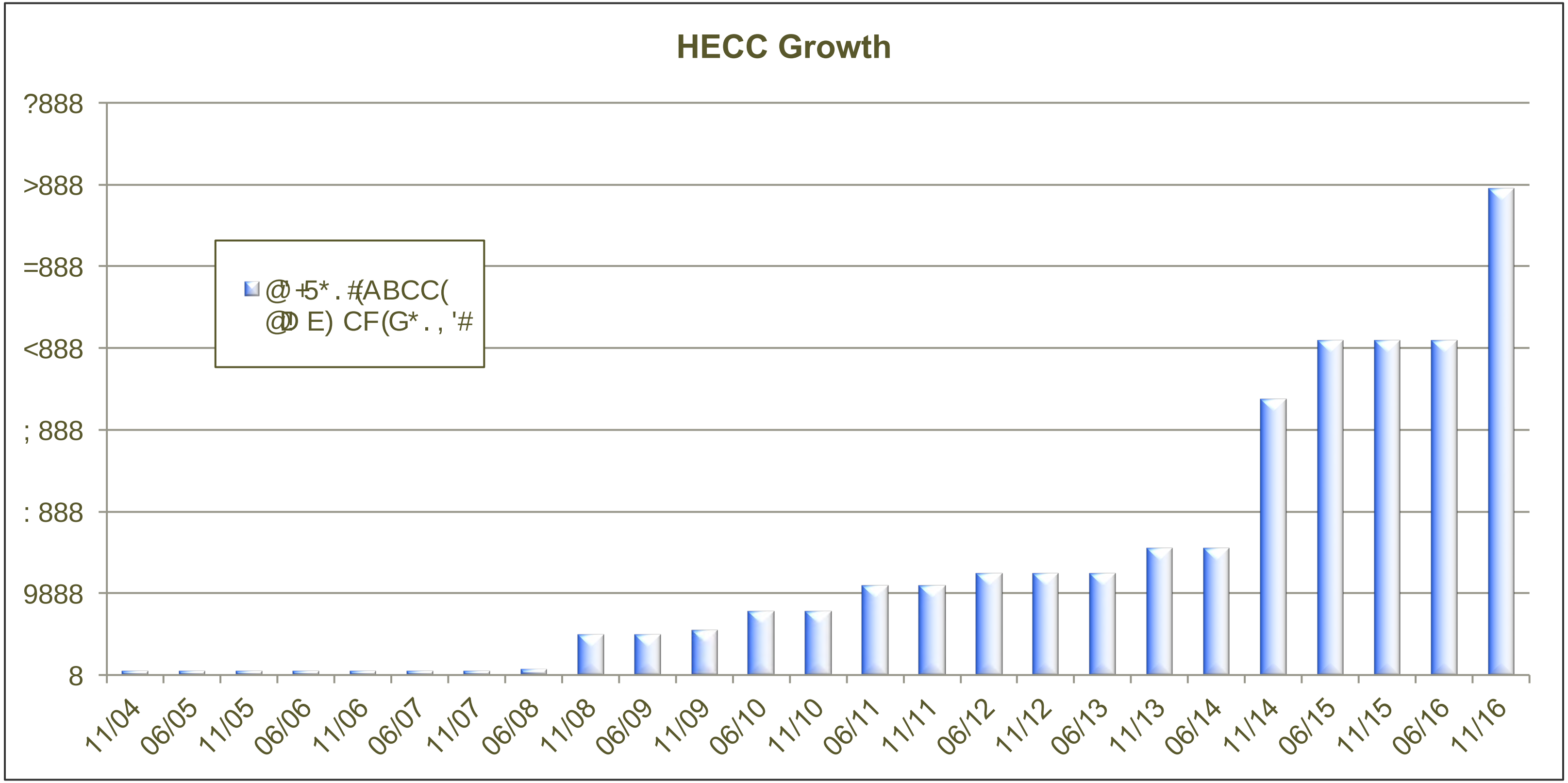
- 3 home file systems 3.7 TB total
- 2 facility-wide scratch file systems 59 TB & 1 PB
- .4 PB for NEX

Archive System

- 490 PB Maximum Capacity
- 6 T950 Spectra Logic Libraries

All storage systems are available to all of the production assets

HECC Growth


HECC Growth
11/04
06/05
11/05
06/06
11/06
06/07
11/07
06/08
11/08
06/09
11/09
06/10
11/10
06/11
11/11
06/12
11/12
06/13
11/13
06/14
11/14
06/15
11/15
06/16
11/16

HECC Services

HECC Conducts Work in Four Major Technical Areas

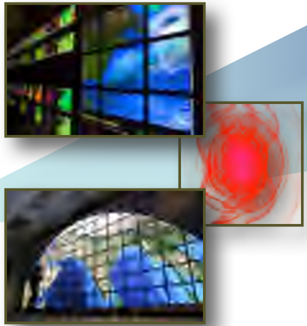
Supercomputing Systems

Provide computational power, mass storage, and user-friendly runtime environment through continuous development of management tools, IT security, systems engineering

HECC

Application Performance and User Productivity

Facilitate advances in science and engineering for NASA programs by enhancing user productivity and code performance of high-end computing applications of interest



Data Analysis and Visualization

Create functional data analysis and visualization software to enhance engineering decision support and scientific discovery by incorporating advanced visualization technologies

Networking

Provide end-to-end high-performance networking analysis and support to meet massive modeling and simulation distribution and access requirements of geographically dispersed users

Supporting Tasks

- **Facility, Plant Engineering, and Operations:** Necessary engineering and facility support to ensure the safety of HECC assets and staff
- **Information Technology Security:** Provide management, operation, monitoring, and safeguards to protect information and IT assets
- **User Services:** Account management and reporting, system monitoring and operations, first-tier 24x7 support
- **Internal Operations:** NASA Division activities hat support and enhance the HECC Project areas

NAS Facility Expansion

Why are We Doing This

The calculation used to be very simple…

- When the cost of maintaining a group of nodes for three years exceeded the cost to replace those nodes with fewer nodes that did the same work, we replaced them.

Now, not so much…

- We look at the total computing our users get and procure new nodes within our budget and remove enough nodes to power and cool the new nodes.
- This means that we are not able to actually realize all of the expansion we are paying for.

But That's Not All

Our computer floor is limited by power and cooling

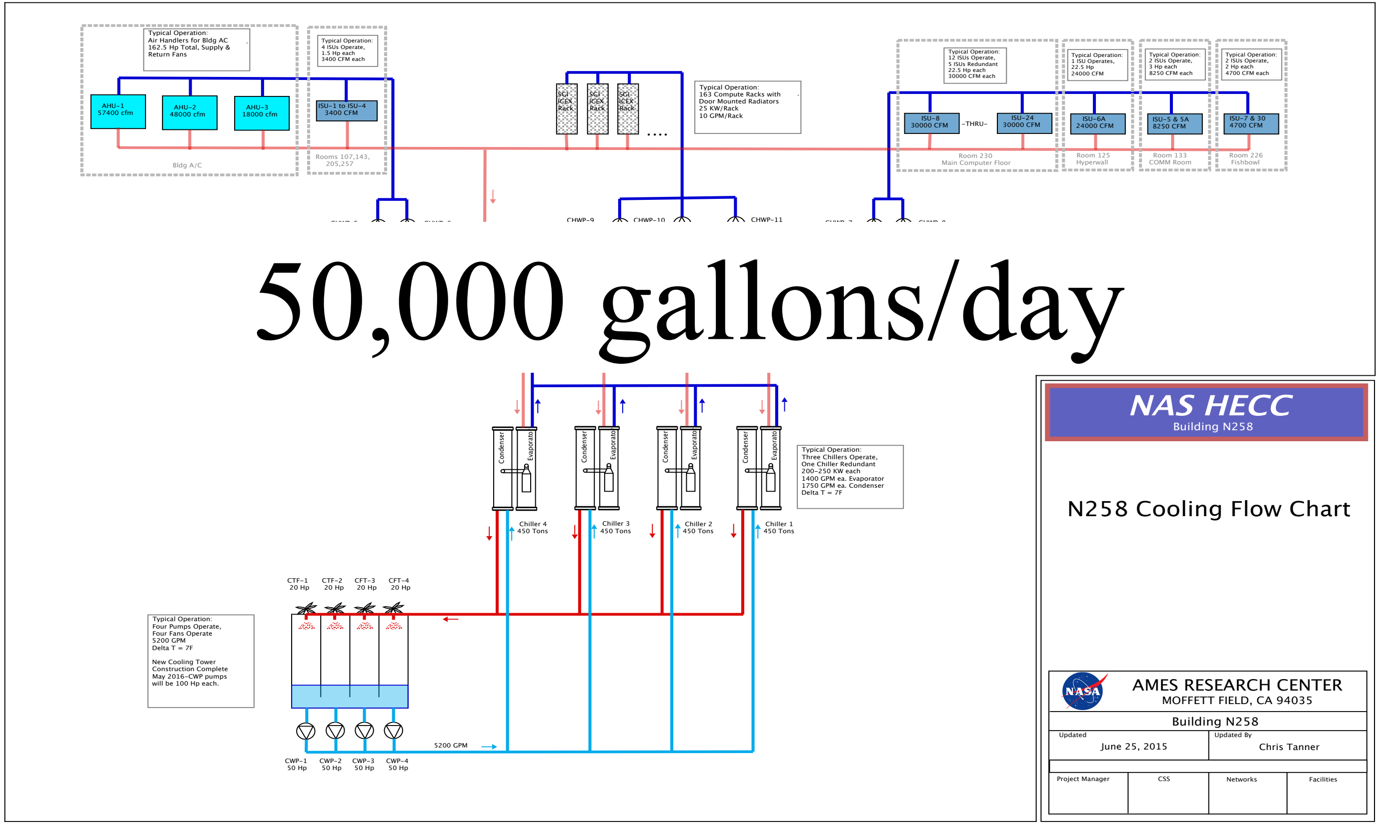
Our Current Cooling System

- 7 Open Air Cooling Tower with 4 50HP pumps
- 7 4 450 Ton Chillers
- 7 7 pumps for outbound chilled water
- 7 4 pumps for inbound warm water

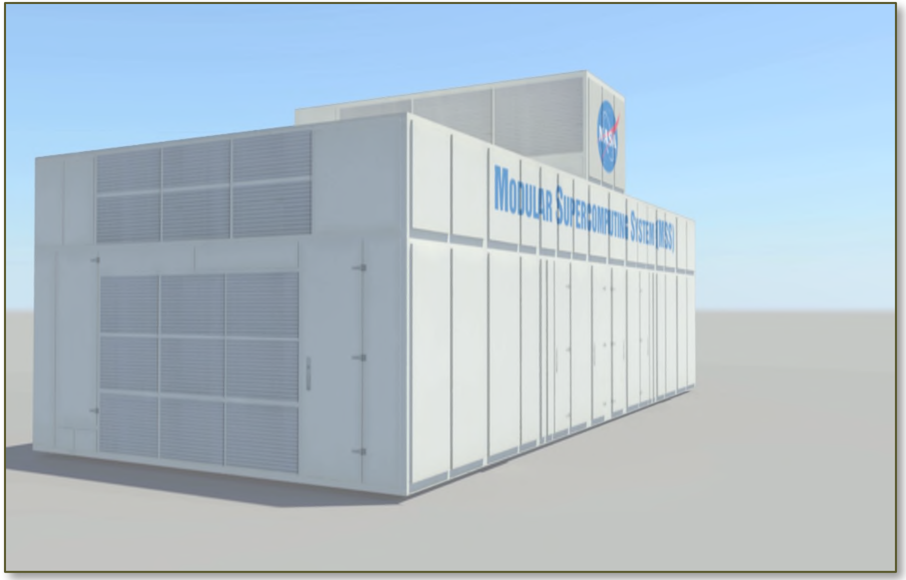
Our Electrical System

- 7 Nominally the facility is limited to 6MW
- 7 20% - 30% is used for cooling
- 7 4MW – 5MW for computing

N258 Cooling Flow Chart



DCU-20



Early Energy Impact

16 Computer Racks (1152 Nodes)	Existing Facility	DCoD-20 Facility	% Savings
H " #* +(I #$$J" #$%&(E* +(K* " +	9L<>8L888(5" 'M	NLN: 8(5" 'MM	OO4<P
B'* - #$ $(Q(1* +(Q* " +	; L; >; LN<8(R3 ST	: LO9=L; : N R3 STT	9NP U%6* +" "V NNP U- %%'$&5V

* Assumes 16 racks represent 8% of facility load
** Adiabatic cooling required 7 days for 5 hours each day
° 1.26 PUE
° ° 1.03 PUE

Progress to Date

We commissioned a study to evaluate the alternatives

We procured and installed prototype modular data center

We received approval to move forward with the full modular center

We're expanding the prototype

We've begun the procurement activities for the full module

Schedule

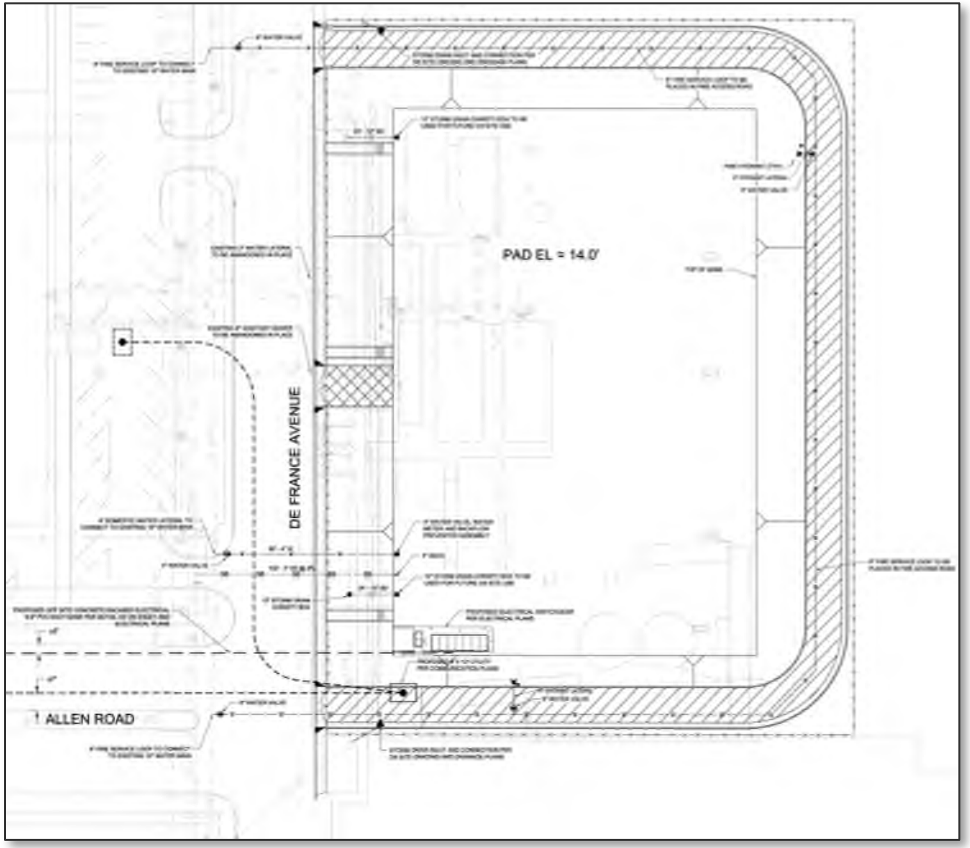
Milestone	Actual/W" +5* #Date(s)
RFI released to vendors	16 January 2017
RFI responses due	15 February 2017
Industry Day at NAS	21 February 2017
C8XY $&, #* (6* &/ %(1+* . * &#" #$%&.	: : (Z* [+, " +Q\ 9(Y " + S(: 89?
G* '* " . * (%](^ +"]#(G* _, * . #[%(E+%1% " '	?(Y " + S(: 89?
G* '* " . * (%](G* _, * . #[%(E+%1% " '	9 Y " Q: 89?
E+%1% " '. (/ , *	; (` , 'Q(: 89?
) 3 " +/ (2 " / *	N(a - #%[* +(: 89?

16-Module Deployment

Site Location

Prepared Site

- 250 ft x 180 ft
- 3 ½ ft of Vertical, Engineered Fill (1½ ft Above DeFrance Road)
- Site Surrounded by DeFrance Rd or Fire Access Road
- Ramp to Top of Elevated Site from DeFrance Road
- 25 kV Switchgear yard at Southwest Corner of Site
- Water Main Point of Connection at Southwest Corner of Site



Prepared Site Utilities

Electrical at Site

25 kV Switchgear Yard (40’ x 12’)

- Four 25 kV Vacuum Circuit Breakers will Distribute up to 15 MVA at 24.9 kV to Step-Down Transformers used in Improved Site
- Power Meters installed on each Vacuum Circuit Breaker
- Site Low Voltage Power
  - 1 Additional 25 kV VCB for Site Power
  - 150 kVA Transformer, 24.9kV/208V, 3 Phase, 4 wire
  - 400 A Panelboard

Water at Site

4-inch Water Main capable of 200 GPM at 40 psi at Point of Connection

- RPZ Backflow Preventer & Water Meter Installed in 4-inch Water Line

Sewer & Storm Drain Piping Installed to edge of Prepared Site

Communications at Site

Data to N258 will be Provided by Conduits & Manholes

Communication Conduits will Terminate at Comm Manhole in Center of Prepared Site

Fiber Optic Procurement & Installation by NASA personnel

Current Approach

- **Our goal is to provide the vendor community the flexibility to propose the solution that best showcases their technology.**
- **The solution must provide the facility expansion to meet the initial deployment with a plan to deploy up to ~18,500 nodes.**
- **NASA's key metric continues to be a measurement of work represented by SBUs. Our current projected SBU growth by fiscal year (Pleiades has 32,578):**
  - FY18 – 15,481
  - FY19 – 15,481
  - FY20 – 18,725
  - FY21 – 18,725
  - FY22 – 22,650
- **NASA will provide the interfaces to the existing file systems providing pairs of MetroX InfiniBand extenders.**
- **There is no storage associated with this procurement.**

Questions

http://www.nas.nasa.gov/hecc