CISL Update
Operations and Services

CISL HPC Advisory Panel Meeting
21 April 2011

Anke Kamrath
anke@ucar.edu
Operations and Services Division
Computational and Information Systems Laboratory

NCAR

Overview

- **Staff Comings and Goings in OSD**
- **Updates:**
  - *HPSS Migration Complete*
  - *NWSC-1 Procurement Update*
  - *NWSC Construction Update*
  - *Restructuring Helpdesk*
  - *VAPOR 2.0 Released*
  - *RDA Updates and Enhancements*
  - *Managing large GAU needs in NSF Proposals*
  - *Storage Allocations (D. Hart)*
  - *Friendly Users for NWSC (D. Hart)*

OSD Staff Comings and Goings...

- **Changes**
  - *Departures*
    - BJ Heller retired January 21, 2011
    - John Merrill retiring May 6, 2011
  - *New & Changed Staff/Positions*
    - *Michele Smart (Allocations/Accounting) moved from ESS to USS*
    - *2 CPG Staff moving to fill in new USS Helpdesk positions*
      - » Scott Baker
      - » Susan Albertson
    - *UCAR Security: Chuck Little*
    - *CISL/NWSC Security: Steve Beatty*
    - *HSS/USS Admin: Linda Yellin*
    - *SE in SSG: Shawn Needham*
    - *SE in DASG/VAPOR: Yannick Polius*
    - *Electrical Lead (Cheyenne): Michael Kercher*
    - *Mechanical Lead (Cheyenne): Jeremy Vaughn*

- **Openings**
  - *1 Documentation/Web*
  - *1 SSG Position (SEII)*

HPSS Migration

- **Completed Migration on March 29, 2011**
- **Went smoothly**
  - *Many user forums/training*
  - *Extensive web documentation*
  - *48 hour outage to:*
    - Dump, translate, reformat, and load the meta-data
    - Reconfigure tape hardware and HPSS software, and test
    - On schedule as planned
  - *MSS meta-data migrated into HPSS*
    - No need to actually "move" data from MSS to HPSS
  - *Many positive user comments on improved performance*

HPSS Migration

- ## What's next
  - *AMSTAR 2 year extension being negotiated*
    - 5 TB per cartridge technology
    - 30 PB capacity increase over 2 years
    - New tape libraries, drives, and media at NWSC in November 2011 for primary copies
    - New tape drives and media at ML in November 2011 for 2nd and Disaster Recovery copies
  - *Planning details of relocation to NWSC*
    - One HPSS system managing primary data copies at NWSC with 2nd and Disaster Recovery copies at ML
    - Migrate existing primary copies to NWSC
    - Utilize 10 GigE link(s) between NWSC and ML
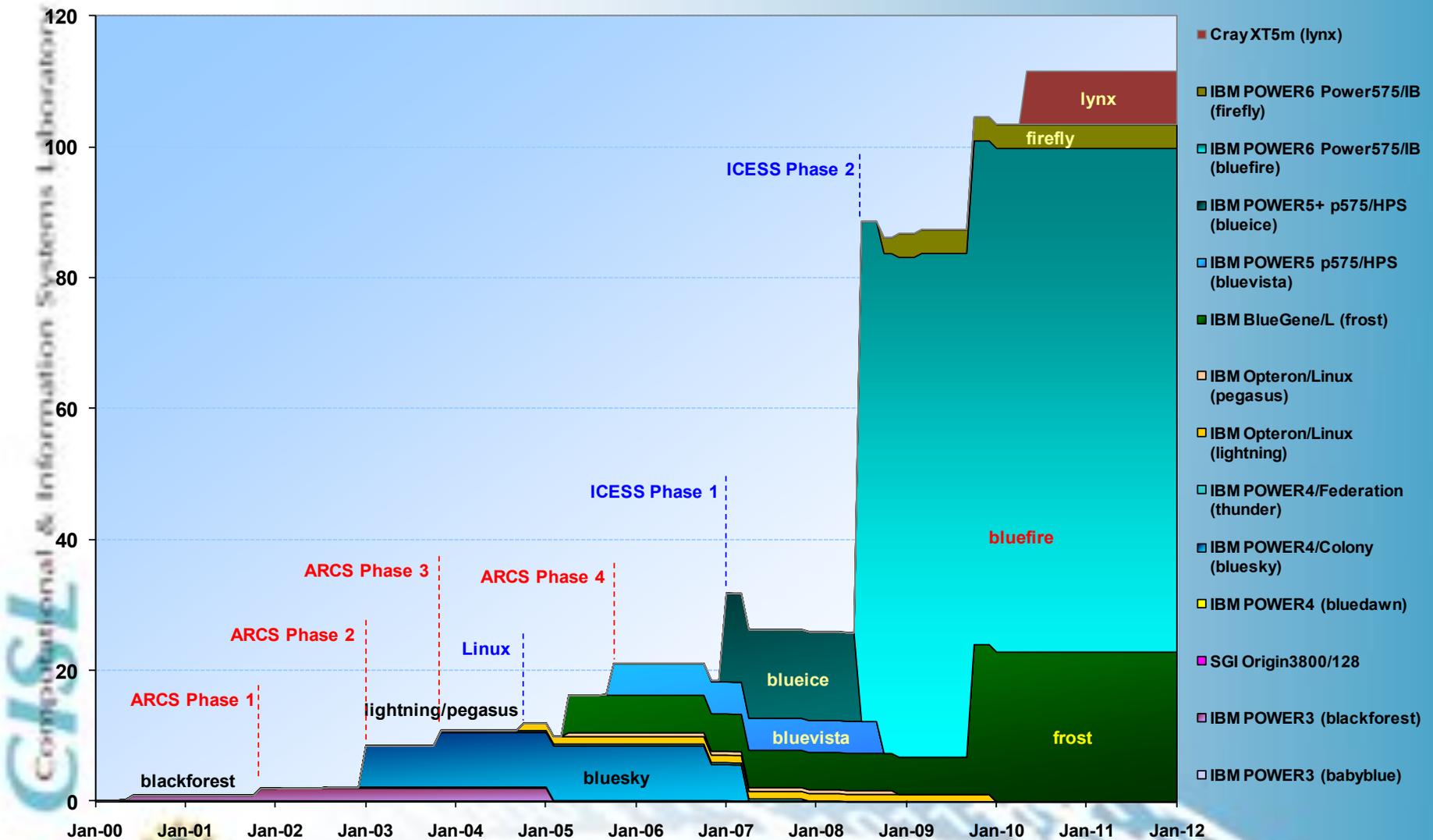
NWSC-1 Procurement Timeline

- **Process began summer 2009**
  - *NWSC HPCT RFI (Fall 2009)*
  - *Initial draft of RFP documents released (Feb 2010)*
  - *SAP input on requirements & benchmarks (Spring 2010)*
  - *TET/BET input on requirements (Summer-Fall 2010)*
  - *TET assistance with benchmarks (Summer-Fall 2010)*
  - *Vendor NDA's (Fall 2010)*

- **NWSC-1 RFP released (17 Dec 2010)**
  - *Mandatory "Vendor Day" @ NCAR (18 Jan 2011)*
  - *Initial proposals received April 5, 2011*
  - *Clarification period & competitive-range down-select*
  - *Final Revised (Best-and-Final) Proposals (request late May; receive mid June)*
  - *Enter negotiations (late July, early Aug)*
  - *Subcontract package to NSF for review/approval (~ 1 Sept)*
  - *Subcontract Award (late September)*
  - *Initial equipment delivery January 2012*
  - *Production Operations mid-2012*

NCAR

HPC Production System(s)

- **One or more systems**
  - *Large number of homogeneous nodes (batch computing)*
  - *High-performance, low-latency interconnect*
  - *Login nodes (≥ 6 nodes for interactive login sessions & submission of batch jobs)*
  - *I/O aggregation nodes*
  - *Connectivity to CFDS resources*
- **Capacity:**
  - *Use NWSC-1 Capacity Benchmarks*
  - *Maximize the total lifetime capacity ('bluefire-years')*
- **Capability:**
  - *Use High-Performance Linpack (HPL) and NWSC-1 Capability Benchmarks*
  - *1Q2012: ≥ 500 TFLOPs with HPL (WY legislative requirement)*
  - *1Q2014: ≥ 1 PFLOPs with HPL*
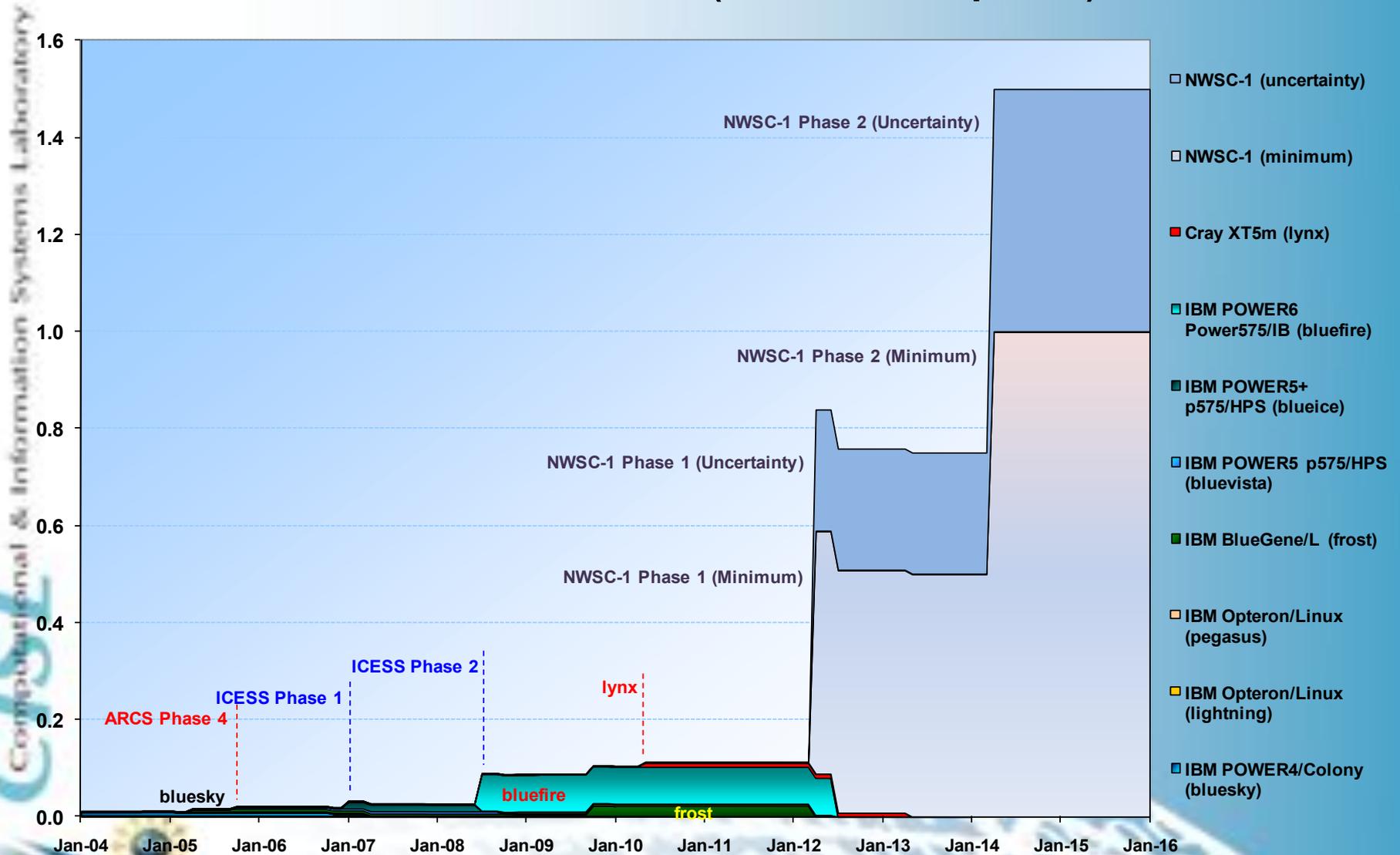- **Request <u>options</u> for expansion, GPU augmentation**

Peak TFLOPs at NCAR (All Systems)

Legend:
- Cray XT5m (lynx)
- IBM POWER6 Power575/IB (firefly)
- IBM POWER6 Power575/IB (bluefire)
- IBM POWER5+ p575/HPS (blueice)
- IBM POWER5 p575/HPS (bluevista)
- IBM BlueGene/L (frost)
- IBM Opteron/Linux (pegasus)
- IBM Opteron/Linux (lightning)
- IBM POWER4/Federation (thunder)
- IBM POWER4/Colony (bluesky)
- IBM POWER4 (bluedawn)
- SGI Origin3800/128
- IBM POWER3 (blackforest)
- IBM POWER3 (babyblue)

Chart labels: ICESS Phase 2, ICESS Phase 1, ARCS Phase 3, ARCS Phase 2, ARCS Phase 4, Linux, ARCS Phase 1, lightning/pegasus, blackforest, bluesky, bluevista, blueice, bluefire, frost, firefly, lynx

Hypothetical scenario 1

Peak PFLOPs at NCAR (NWSC-1 two phase)

Peak PFLOPs at NCAR (NWSC-1 single drop)

Chart: Peak PFLOPs at NCAR plotted from Jan-04 to Jan-16 (x-axis) versus 0.0 to 1.6 (y-axis). Labeled regions include ARCS Phase 4, ICESS Phase 1, ICESS Phase 2, lynx, bluesky, bluefire, frost, NWSC-1 Phase 1 (Uncertainty), and NWSC-1 Phase 1 (Minimum).

Legend:
- NWSC-1 (uncertainty)
- NWSC-1 (minimum)
- Cray XT5m (lynx)
- IBM POWER6 Power575/IB (bluefire)
- IBM POWER5+ p575/HPS (blueice)
- IBM POWER5 p575/HPS (bluevista)
- IBM BlueGene/L (frost)
- IBM Opteron/Linux (pegasus)
- IBM Opteron/Linux (lightning)
- IBM POWER4/Colony (bluesky)

CFDS Production Systems

- **One or more systems**
  - *Filesystems (software)*
  - *Filesystems servers and Data Storage resources*
  - *High-performance external connectivity (e.g. InfiniBand)*
  - *On-site spare parts*

- **Capacity**
  - *Prototype filesystem allocation:*
    - */scratch ~50%, /projects ~35%, /users ~15%*
  - *1Q2012: ≥ 6 PB usable*
  - *1Q2014: ≥ 15 PB usable*

- **Capability**
  - *1Q2012: I/O burst write ≥ 75 GB/sec, sustainable read/ write rate ≥ 30 GB/sec for the two largest filesystems, "burst" is 20% of HPC aggregate memory; or ~20 TB*
  - *1Q2014: I/O burst write ≥ 150 GB/sec, sustainable read/ write rate ≥ 60 GB/sec for the two largest filesystems, "burst" is 20% of HPC aggregate memory; or ~40 TB*

- **Request <u>options</u> for expansion**

DAV Production Systems

- **One or more systems (Intel x86_64 instruction set, w/ CUDA, OpenGL & OpenCL, graphics cards capable of > 1 TFLOP)**
- **1Q2012**
  - *Large Memory Nodes*
    - 512 cores, 10 TB total memory or more
      ("two 1 TB memory jobs + twenty ≤ 512 GB memory jobs")
    - 60 GB/s aggregate (4 GB/s single-stream) IO to CFDS
    - 1 graphics card/node, or 8 graphics cards, whichever larger
  - *GPU-Computation/Visualization Cluster*
    - Sixteen nodes each with 64 GB memory, at least 8 cores/node
    - 40 GB/s aggregate (4 GB/s single-stream) IO to CFDS
    - At least 1 graphics card per CPU socket
    - Drive Vis-wall
- **1Q2014**
  - *Request <u>option to</u> ~double the above*
- **Trend: More NCAR-centric DAV efforts due to size of data. Processing 100s TB on university resources challenging and costly.**

NWSC Construction Update

- **All Major Construction Components are Delivered and Installed**
- **Permanent Electrical Power**
  - *Energized 24.9 KV equipment April 6th*
- **Mechanical Systems Startups**
  - *Heating water loops May 5th*
  - *Chilled water loops May 19th*
  - *Air handling unites June 1st*
- **Functional Testing & Systems Testing**
  - *June – August*
- **Building is on track to be substantially complete by early August**
- **Will initiate full Integrated System Testing**
 - *August*

Restructuring Help Desk

- **Help desk function being moved from Operations (CPG) to User Services**
  - *2 staff are moving from CPG in May 2011*

- **In support of operational changes for NWSC**
  - *Operations staff will be more system focused and move to Cheyenne*
  - *Help desk will remain at Mesa Lab*

- **Changes**
  - *Help desk to provide more technical HPC support*
  - *Help desk will support user documentation and other web publications*
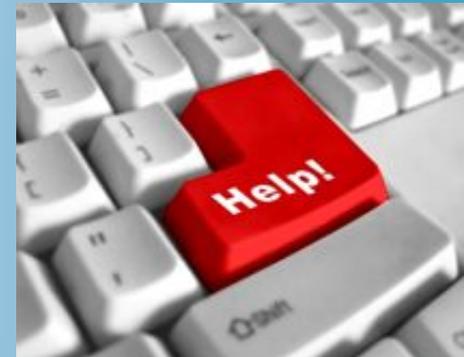
VAPOR 2.0 Released
Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers

- **http://www.vapor.ucar.edu/**

- **Features:**
  - *Increased Python Support*
  - *Data Compression*
  - *Direct import of WRF-ARW output files*
  - *Improved User Interface*
  - *Faster Rendering of Flow Lines*
  - *Native Mac OSX Support*
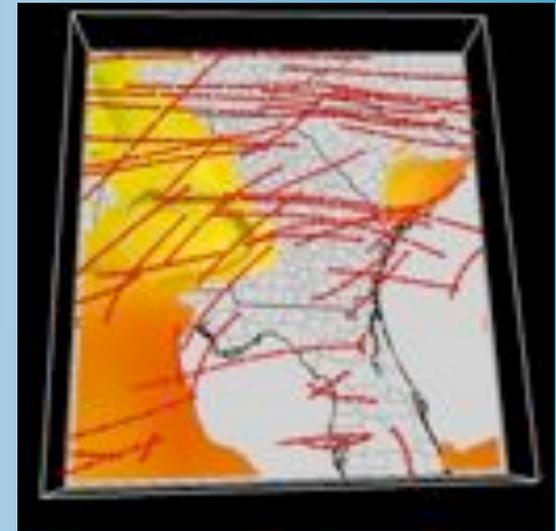
RDA, ECMWF Recent and Future Enhancements

Enabled by client driven access to ECMWF mass storage system & saving $8-10K annually

ECMWF Re-analysis Interim (ERA-I)

- Resolutions: 512x256, 6-hourly
- Time Period: 1989 – Jan. 2011, updated quarterly

Year of Tropical Cyclone (YOTC) Dataset

- Resolutions: T799, 6-hourly
- Time period: May 2008 – May 2010

High Resolution Operational Analysis (future)

- Resolutions: T1279, 6-hourly
- Time period: Jan 2010 - ongoing

Managing Large GAU Needs in NSF Proposals

- **NSF concerned that 5x oversubscription may mean sub-critical amount of GAUs to support proposals**
- **Should there be a pre-CHAP request for needs above 600K (now) or 5M (NWSC) GAUs?**
  - *How would this work?*
- **Is it necessary?**
  - *Proposers can come back and ask for more.*
  - *There have been no complaints to NSF*
- **Right-sizing compute to fit programmatic activities**
  - *NSF contributing funds for compute to support EaSM*

Questions?

CHAP Meeting
21 April 2011

NCAR