

NCCS User Forum

July 19, 2011



Agenda



- Introduction (Lynn Parnell, NCCS HPC Lead)
- Discover Update
- Archive Update
- User Services Update
- Data Services & Analysis Update
- New Initiatives
- Question & Answer

- Breakouts...



Post Meeting Breakouts



- Breakouts:
 - Ultra-Scale Climate Data Analysis Tools (Tom Maxwell)
 - SIVO PyD, a Python Distribution for Scientific Data Analysis (Jules Kouatchou)
 - NCCS Twitter Feed (Jarrett Cohen)



Discover Update

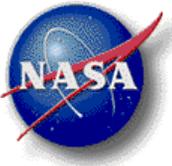
Dan Duffy, NCCS Lead Architect
and
Fred Reitz, NCCS Operations Lead



Discover SCU7 – Westmere Nodes



- 40% have been dedicated to large scale application/research runs.
- In the process of opening access up for general use.
- Currently running general-access SCU7 queues with same names as “lower Discover” queues (but a separate PBS domain).
- Existing specialized SCU7 queues:
 - CMIP5/IPCC AR5 queues will remain.
 - Large core count (>1000) queues will remain but will not have current priority configuration.
 - Possible queue modifications as workload demands change.



Discover Upgrade to SLES11 and PBS10



- Plan to upgrade as many nodes as possible to SLES11 and PBS 10.
 - SCU7 currently running this software stack.
 - SCU3 and SCU4 will use this software stack after the hardware upgrade.
- Must phase this across the other nodes.
 - SCU5 and SCU6
 - Dali Analysis Nodes
- Note that the Base, SCU1, and SCU2 are not going to be upgraded at this time.
 - The NCCS is investigating hardware options to upgrade these nodes.
- As with any O/S upgrade, some user changes may be required.
 - Best practice to recompile your applications.
- NCCS web site will have documentation on user changes needed for SLES11 and PBS 10.



Upgrade to SCU3 and SCU4



- The NCCS plans on swapping out the Harpertown nodes for Westmere nodes.
- Includes swapping out the compute nodes and chassis within SCU3 and SCU4 for new hardware.
- Reuse of all the infrastructure (such as racks, power, cooling, switching, etc.).
- Significant overall change of
 - Harpertown 2.5 GHz to Westmere 2.8 GHz processors
 - Gain of 2K cores and about 36 TF peak
- The current plan (subject to change) is to begin these modifications as early as next week.
 - The target is to have the processors back up for general use by the end of August.



Discover GPU Systems Available



- 32 compute nodes each configured with
 - Dual socket, hex-core Intel Westmere 2.8 GHz processors
 - 48 GB of RAM
 - QDR Infiniband
 - All GPFS file systems
 - Two Nvidia M2070 GPUs
 - Each with 488 CUDA cores and 6 GB memory (28,672 Total GPGPU Streaming Cores)
- PGI CUDA FORTRAN compiler for code-porting
- Discover GPUs available via ‘warp’ queue
 - Accessed through the SLES11 nodes at this time
 - Send email to support@nccs.nasa.gov to request access



NVIDIA M2070 “Tesla” GPGPU



Discover Scalable Units After SCU3 and SCU4 Upgrades



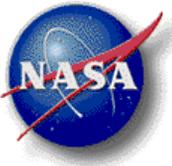
SCU	Year	Vendor	Processor	Cores	TF
Base	4Q 2006	LNXI/Supermicro	3.2 GHz Intel Xeon Dempsey (2 flop/s per clock)	520	3.33
1	2Q 2007	LNXI/Dell	2.66 GHz Intel Xeon Woodcrest (4 flop/s per clock)	1,032	10.98
2	3Q 2007	LNXI/Dell	2.66 GHz Intel Xeon Woodcrest (4 flop/s per clock)	1,032	10.98
3	3Q 2011	IBM	2.8 GHz Intel Xeon Westmere (4 flop/s per clock)	3,096	33.86
4	3Q 2011	IBM	2.8 GHz Intel Xeon Westmere (4 flop/s per clock)	3,096	33.86
5	3Q 2009	IBM	2.8 GHz Intel Xeon Nehalem (4 flop/s per clock)	4,128	46.23
6	1Q 2010	IBM	2.8 GHz Intel Xeon Nehalem (4 flop/s per clock)	4,128	46.23
7	4Q 2010	Dell	2.8 GHz Intel Xeon Westmere (4 flop/s per clock)	14,400	157.5
Totals				31,432	342.97



Other Major Changes



- Addition of approximately 900 TB of DDN disk to be added into the GPFS file system.
 - The disks are currently in house and being tested.
- Other?



Discover Linux Cluster: Other



- Reduced incidence of GPFS hangs and nodes run out-of-memory—more automated monitoring.
- Coming ‘login’ bastion host changes:
 - Dali would be accessible via ‘login’ only.
 - Please describe the impact of this change via online survey, available through July 23:
https://www.nccs.nasa.gov/user_admin_forms/bastion-login-survey.php



Archive Update

Tom Schardt, NCCS Archive Lead



Archive: Upgrades



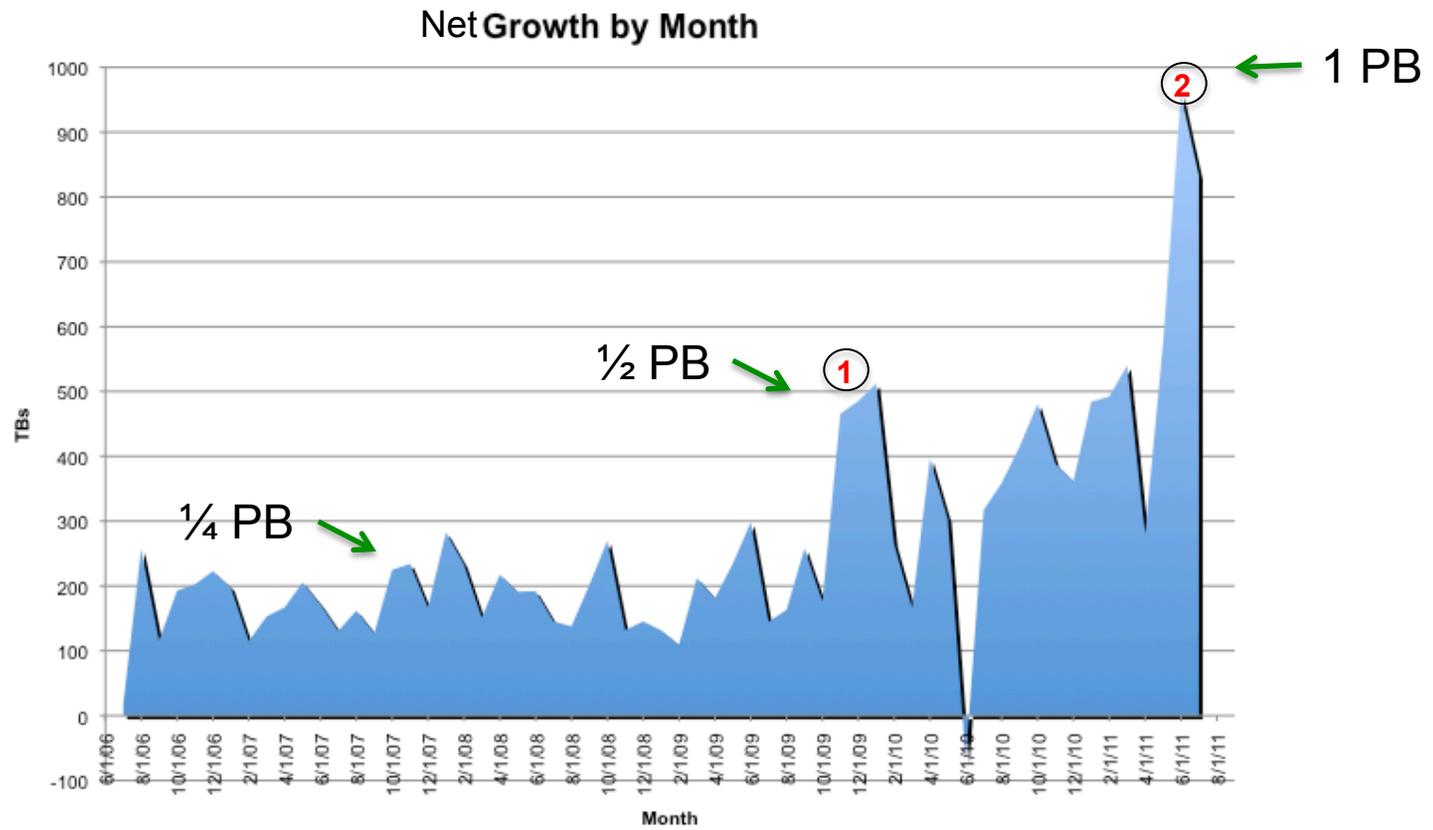
- 442 TBs of additional disk cache (adds ~80%)
 - Users will be contacted before their archive directories are moved.
- Tape Library Upgrades
 - Oracle T10000C tape drives – 5 TB per tape.
 - Replacement tape library for building 32 second copies
 - Requires moving all existing second copies to the new tape media.
 - Please delete files you do not need so they are not moved; thus decreasing the time required to move data to the new media and freeing tape drives for recalling user data.



Archive: Growth



- Current holdings are 19.2 PBs
- Significant growth rate increases in 10/2009 and 5/2011

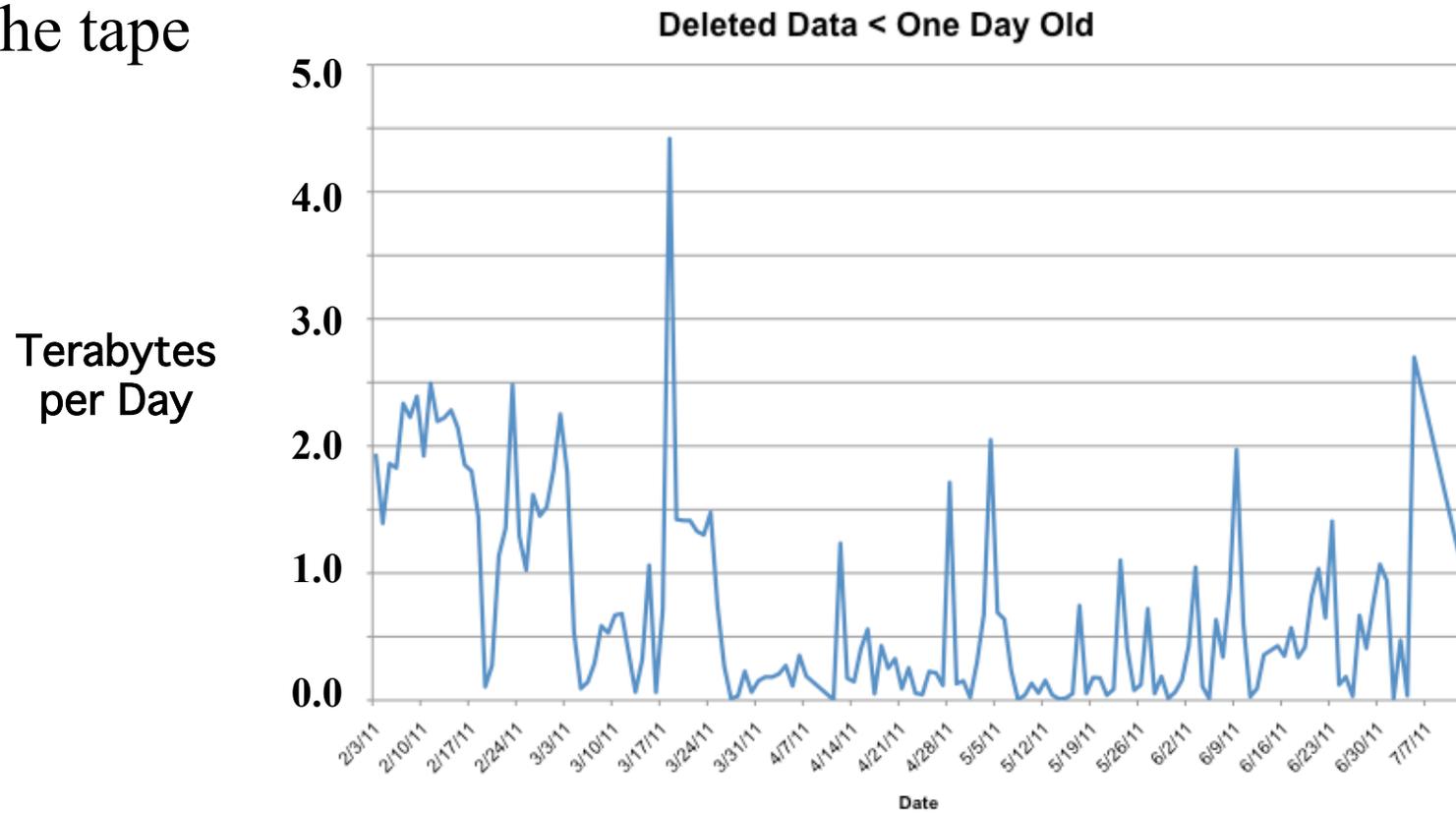




Archive: Growth



- Some users are using the Archive as scratch space
- To reclaim tape space of deleted files requires rewriting the tape



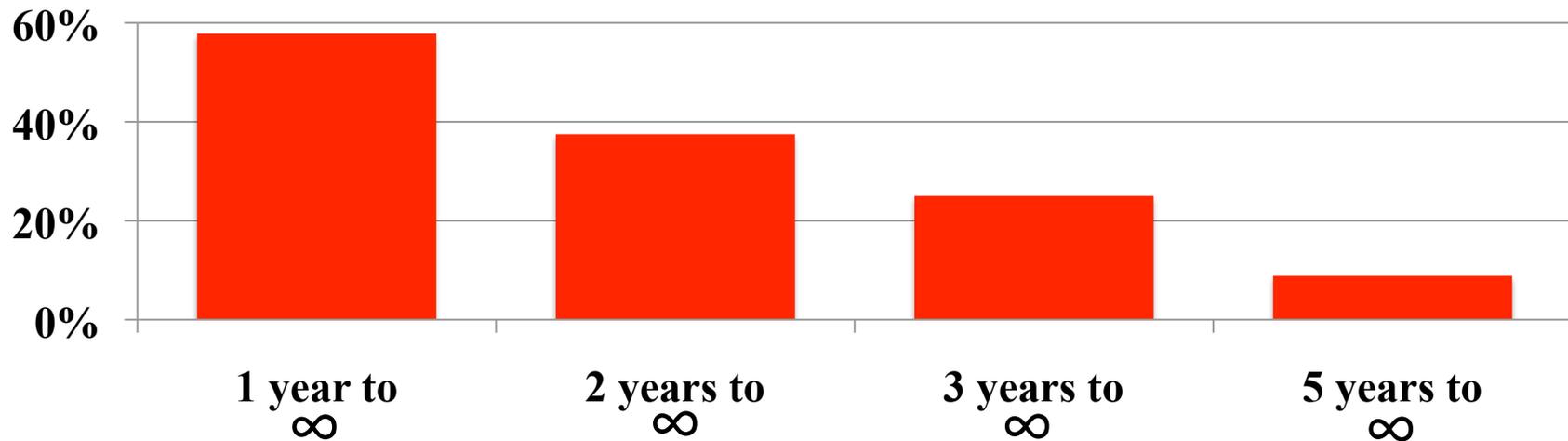


Archive: “Old” Data



Last Access Time	Amount of DMF Data
1 to 2 years	3.9 PBs
2 to 3 years	2.4 PBs
3 to 5 years	3.1 PBs
> 5 years	1.7 PBs

Percentage of Data Last Accessed This Long Ago *or Longer*

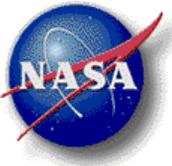




Archive Growth and Policy



- NCCS is developing plans for managing Archive data growth.
- We'll be soliciting NCCS user and stakeholder input.
- Stay tuned for future announcements...



User Services Update

David Kistler, User Services Manager



User Services – Staff Additions



- David Kistler
- George Fekete



New RSA Tokens Coming



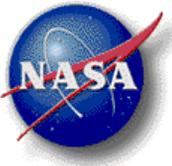
- NCCS RSA tokens will be replaced.
- More details on replacement logistics when we get them from the Agency.



User Services: “Musty Ticket” Cleanup



- Upcoming Housekeeping Effort
 - Purpose: to clear out old/OBEEd tickets from the FootPrints database
 - USG will comb through the NCCS Ticket Manager workspace in FootPrints
 - USG will identify candidate tickets for closure
 - USG will contact the originator of each candidate ticket to get permission to close
 - USG will subsequently close the ticket



User Services Group – Contact Information



- NCCS User Services:
 - support@nccs.nasa.gov
 - 301-286-9120
- NCCS User Services Manager:
 - Dave Kistler
 - Email: david.m.kistler@nasa.gov
 - Phone: 301-286-9925



Data Services & Analysis Updates

Laura Carriere, NCCS Data Services
and
Tom Maxwell, NCCS Analysis Lead



Data Services & Analysis



- Working on logistics for export to NCCS VisWall system of specific requested Discover data which owners declare to be “public”, please send requests to support@nccs.nasa.gov
- Preview of Tom Maxwell’s Ultra Scale Climate Data Analysis Tools breakout

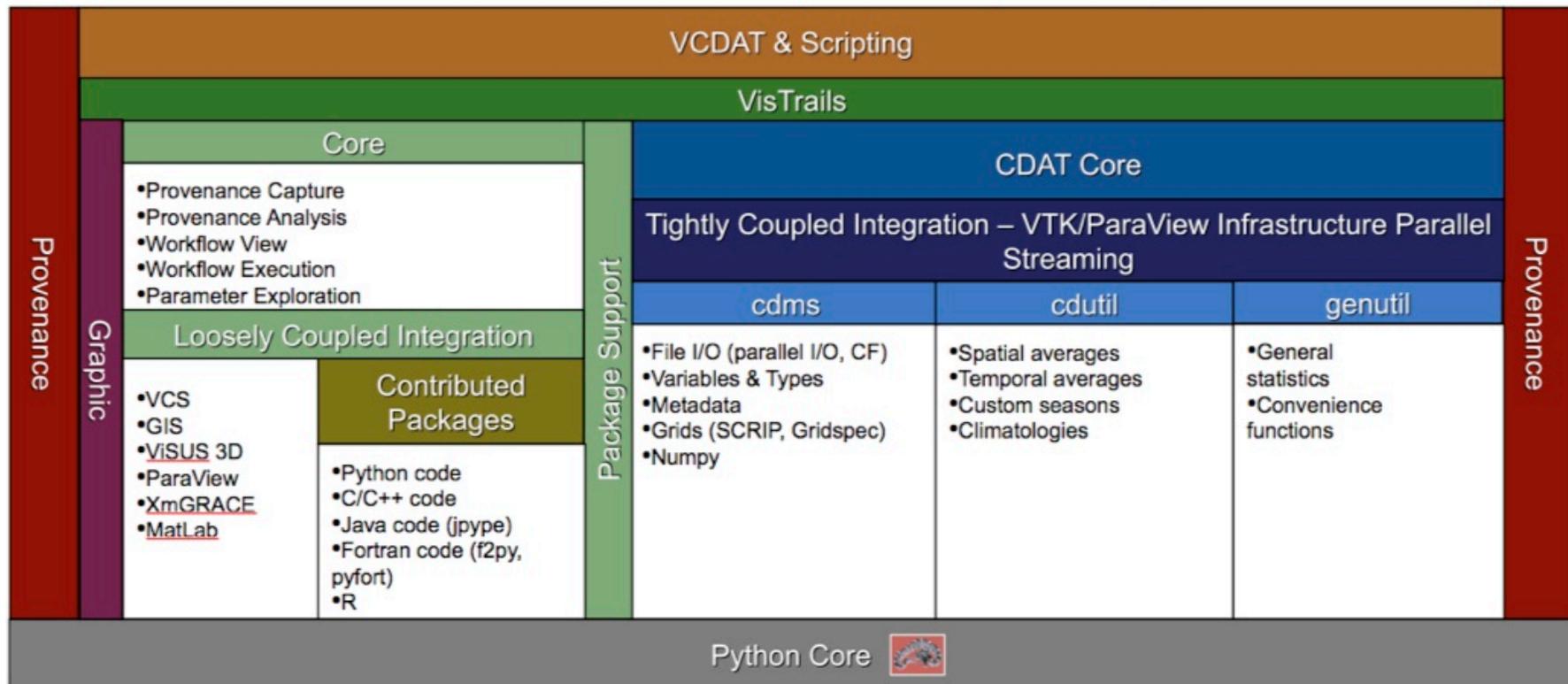


High Performance Climate Computing



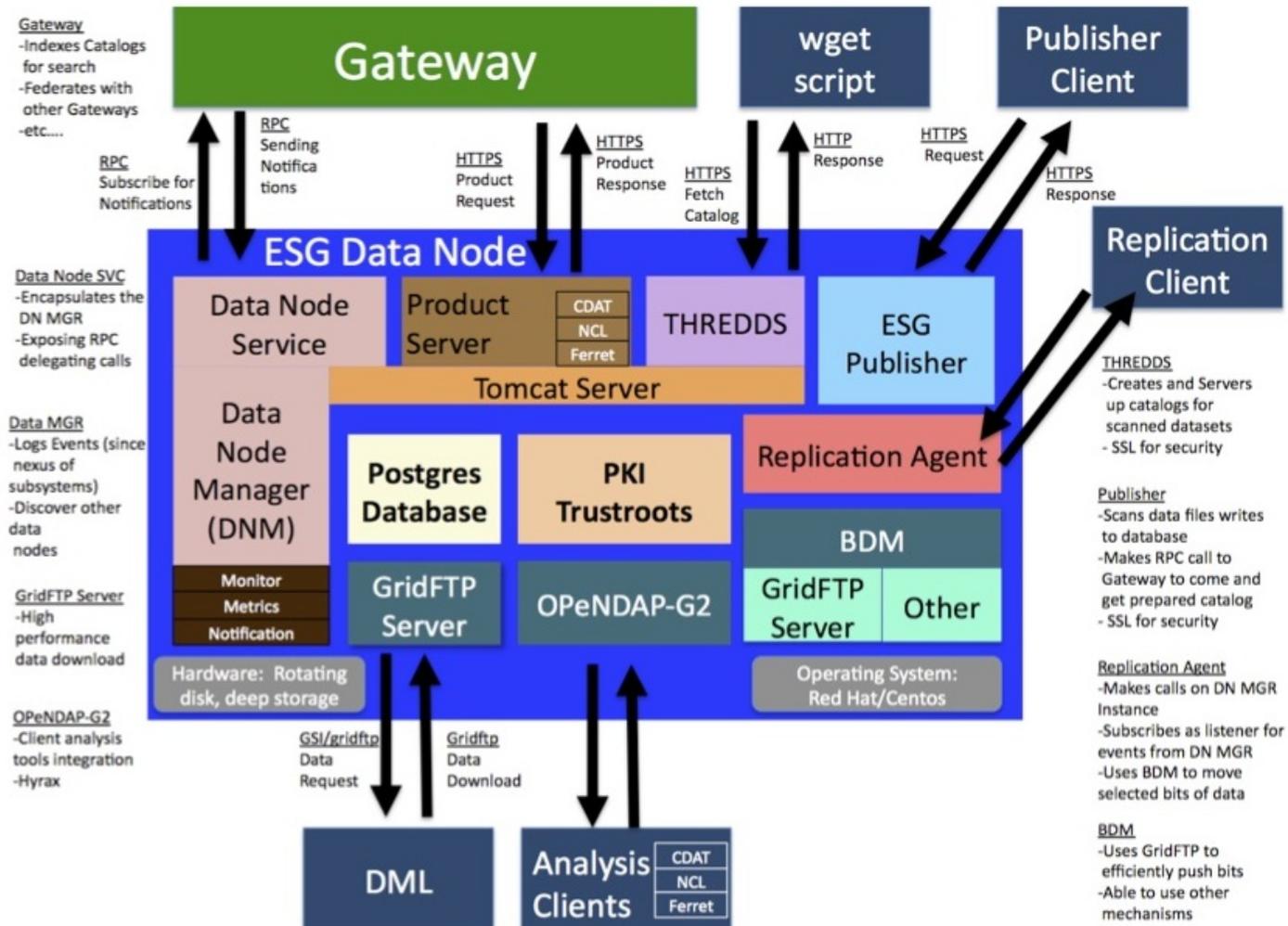
- ❑ Many legacy FORTRAN and Matlab codes will be ported to HPC.
- ❑ Build upon a set of commonly used climate data processing primitives.
- ❑ Requires a standard climate data processing environment.

Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT) Architectural Layers





Climate Data Analysis in ESG





Climate Data Visualization and Analysis with UV-CDAT/DV3D



The screenshot displays the VisTrails interface. The main window, titled "VisTrails Builder - DemoWorkflow9.vt*", shows a workflow diagram with the following components:

- CDMS_FileReader** (yellow box) feeds into a **Difference** module (green box).
- The **Difference** module feeds into three **CDMS_VolumeReader** modules (green boxes).
- The first **CDMS_VolumeReader** feeds into a **VolumeRenderer** (green box), which then feeds into a **DV3DCell** (green box).
- The second **CDMS_VolumeReader** feeds into a **VolumeSlicer** (green box), which then feeds into a **DV3DCell** (green box).
- The third **CDMS_VolumeReader** feeds into a **LevelSurface** (green box), which then feeds into a **DV3DCell** (green box).

The right-hand window, titled "VisTrails - Spreadsheet - Untitled", shows the configuration for the **CDMS_FileReader** module. The **Methods** panel lists the following signature:

Method	Signature
CDMS_FileReader	
datasetId	(String,Integer)
datasets	(String)
grid	(String)
roi	(Float,Float,Float,Float)
timeRange	(Integer,Integer,Float,Float)

The **Set Methods** panel shows the following configuration:

- datasets**: String eloper/Data/AConaty/comp-ECMWF/ac-comp1-geos5.xml
- datasetId**: String ac-comp1-ecmwf
- Integer**: 10

The **CDMS_FileReader Module Configuration** panel shows the following configuration:

- Dataset**: ac-comp1-ecmwf
- Buttons: Add Dataset, Remove Dataset, OK, Cancel

The bottom-left **Console** window shows the following output:

```
--- Set Range: (-107.356263, 85.754088), Initial Range = ( 0.000000, 2.000000 ), P = ( 221, 25 ) dP = ( -1.318182, 1.038136 )
setLevelRange, data range = [-107.3562632415254, 85.754088320974617, 0]
Update 1 Level(s), range = [-3828.800515, 18512.676757 ], levels = [0.0, 7341.938120875084]
Update levelRangeScale Leveling, data = [-107.3562632415254, 85.754088320974617, 0]
--- Set Range: (-113.356867, 79.753484), Initial Range = ( 0.000000, 2.000000 ), P = ( 221, 24 ) dP = ( -1.318182, 1.059322 )
setLevelRange, data range = [-113.35686738612291, 79.753484176377114, 0]
Update 1 Level(s), range = [-4523.027210, 17818.450062 ], levels = [0.0, 6647.711425959822]
Update levelRangeScale Leveling, data = [-113.35686738612291, 79.753484176377114, 0]
PM_LevelSurface.Persist-Parameter-List[ac-comp1-ecmwf.Height-hght-Difference.] (v. 241): [['levelRangeScale', [-113.35686738612291, 79.753484176377114, 0]]]
process Key Event, key = SHIFT_L
process Key Event, key = SHIFT_L
-- Key Press:
```



Vistrails Provenance Management



Vistrails Builder - lung.xml

Search & Refine

Query Template

Search Refine Reset

Properties

Version Tag Change

User

Date

Notes

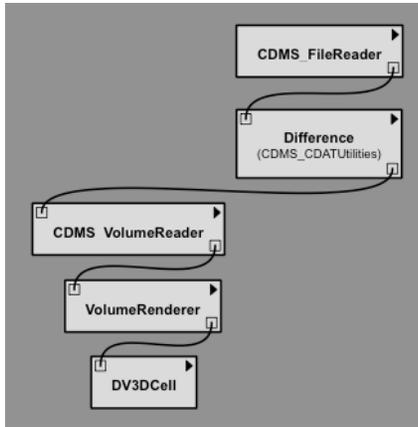
Versions created by the current user are blue. Versions created by all other users are orange.



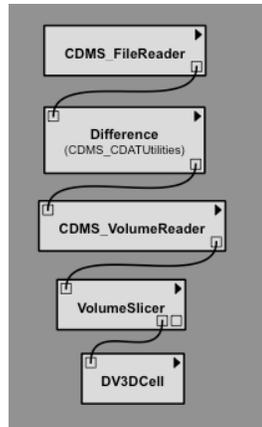
Parallelism in Climate Data Analysis



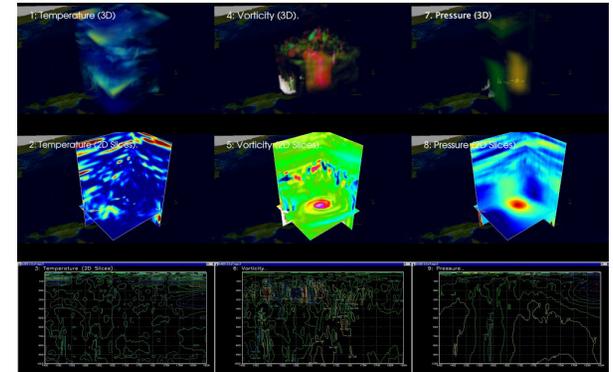
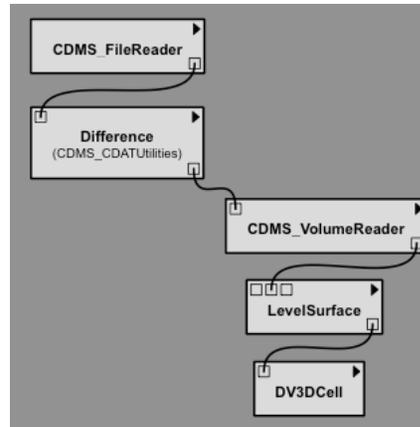
Client-0



Client-1

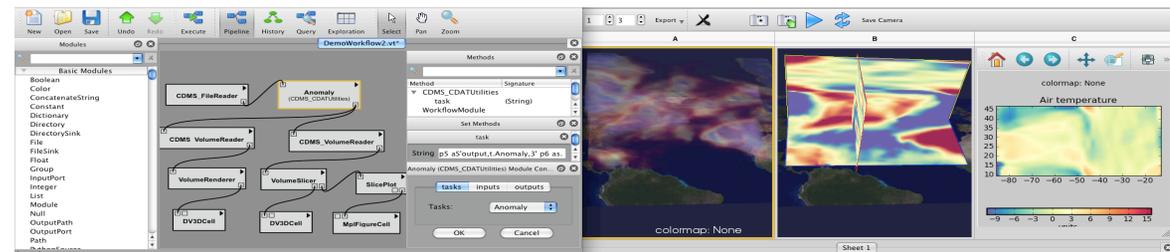
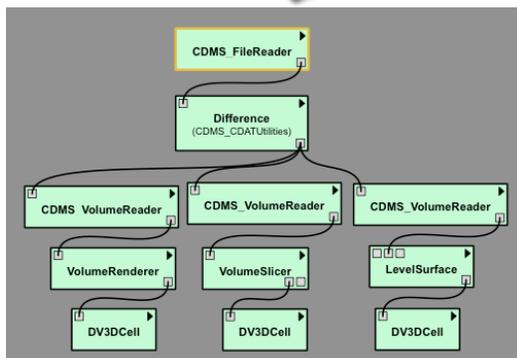


Client-2



- Data Parallelism
- Task Parallelism
- Time Parallelism

Server





New Initiatives

Dan Duffy,
NCCS Lead Architect



New Initiatives



- Will be forming an “NCCS User Advocacy Board” to expand opportunity for user stakeholder input on NCCS plans & policies.



Questions & Answers

NCCS User Services:

support@nccs.nasa.gov

301-286-9120

<https://www.nccs.nasa.gov>



Contact Information



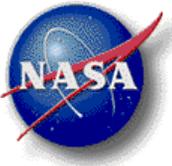
NCCS User Services:

support@nccs.nasa.gov

301-286-9120

<https://www.nccs.nasa.gov>

Thank you



Breakout:

Ultra-Scale Climate Data
Analysis Tools

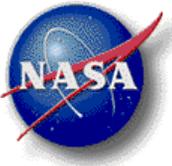
Tom Maxwell



Breakout:

SIVO PyD,
a Python Distribution for Scientific Data
Analysis

Jules Kouatchou



SIVO-PyD

A Python Distribution for Scientific Data Analysis

Jules Kouatchou

SIVO Code 610.3

Jules.Kouatchou-1@nasa.gov



Background (1)



- Python is used in industry and academia for a wide variety of purposes
- There are more than 15,000 Python related packages
- Few attempts to regroup some of them in a distribution for scientific computing
- Available products are mainly commercial



Background (2)



- GMAO adopted Python for model development and data analysis
- Other NCCS users rely on Python
- In 2009, we started the process of creating a Python distribution



Background (3)



- Parts of the distribution were installed by different staff
- The distribution was static (not possible to change the version of Python)
- Users' requests were not handled in a consistent way

New Approach: **SIVO-PyD**



What is SIVO-PyD?



- Collection of Python packages for scientific computing and visualization
- All the packages are accessible within the Python framework
- Self-contained distribution that mimics the commercial Enthought one.



Design Philosophy



- Only use open-source packages
- Follows baselibs concepts
- Automatically installed on *discover* and on **Mac**
- Does not alter source codes
- New packages or new versions of existing ones are added with minimal effort
- Maintainable in a version control repository



SIVO-PyD Main Packages



Python

Numpy

ipython

pygrads

pyhdf

matplotlib

SciPy

h5py

netCDF4



How to Use the Distribution



The distribution is available to any NCCS user on *discover/dali*.
You only need to load the modules:

`other/comp/gcc-4.5`

`other/SIVO-PyD/spd`



More Information



<https://modelingguru.nasa.gov/clearspace/docs/DOC-2109>

In case you want additions to be made to the distribution, contact:

Jules Kouatchou

Jules.Kouatchou-1@nasa.gov



Breakout:

NCCS Twitter

Jarrett Cohen



NASA and



- Twitter is a social media platform with approximately 200 million users worldwide.
- “Tweets” are short messages of at most 140 characters. They sometimes include a clickable web link to provide more information.
- While Twitter is well-known for celebrity use, many businesses and organizations use Twitter to communicate with customers and enthusiasts.
- NASA is particularly well regarded for its use of social media. A wide variety of NASA Twitter feeds are accessible from: <http://www.nasa.gov/connect/index.html>
- The main NASA feed has more than 1.2 million “followers,” Twitter account holders who subscribe to the feed.





“NCCS User News” Twitter Feed



- The “NCCS User News” Twitter feed provides existing Twitter users and others a short-form option for accessing NCCS information.
- NCCS tweets include **system status information**, **hardware/software updates**, **events** (e.g., user telecons, user forums), and **tips** (e.g., from Modeling Guru). We also post occasional links to **NCCS news releases and stories** as well as links to **relevant items from computing and science publications** (e.g., *HPCwire*).
- Information provided in User Services Group emails will be available in summary form on Twitter, with links to the NCCS website for additional details.
- General Twitter Access
 - Twitter is accessible through web browsers, desktop software (e.g., TweetDeck), and smartphone apps.
- “NCCS User News” Access
 - Anyone with a web browser: view the feed at http://twitter.com/NASA_NCCS
 - Twitter account holders: search for and follow “@NASA_NCCS”
 - Note: For at least the next few weeks, “NCCS User News” is a private feed and will require signing up for a Twitter account (if you don’t already have one) and requesting access. Someone from our Twitter team will then approve access.





“NCCS User News” Preview



NCCS User News

@NASA_NCCS Greenbelt, MD USA

NASA Center for Climate Simulation (NCCS) system status, hardware/software updates, events, tips, and more. Contact us at 301-286-9120 or support@nccs.nasa.gov.
<http://www.nccs.nasa.gov/>

Following

Tweets

Favorites Following Followers Lists



NASA_NCCS NCCS User News

NCCS Weekly User Teleconference today, Tues., Jul. 12 at 1:30 PM ET. <http://bit.ly/ogYNe5>
22 hours ago



NASA_NCCS NCCS User News

Submissions due Sept. 4 for Computing in Science & Engineering special issue on GPU applications. <http://bit.ly/r9IWJC> (via [@HPC_Guru](https://twitter.com/HPC_Guru))
12 Jul



NASA_NCCS NCCS User News

Join us at the NCCS User Forum next Tues., Jul. 19 at 2:00 PM ET. <http://bit.ly/p8OCPP>
11 Jul



NASA_NCCS NCCS User News

RT [@HPCwire](https://twitter.com/HPCwire): Around the Web: Outside the Climate Change Model Box <http://bit.ly/nwUJgb> #HPC
7 Jul



NASA_NCCS NCCS User News

Reminder: User survey on enhancing bastion login environment. We would appreciate your input. <http://bit.ly/jHudrM>



About @NASA_NCCS

16

Tweets

0

Following

2

Followers

0

Listed

Following · [view all](#)

[About](#) [Help](#) [Blog](#) [Mobile](#) [Status](#) [Jobs](#) [Terms](#) [Privacy](#)
[Shortcuts](#) [Advertisers](#) [Businesses](#) [Media](#) [Developers](#)
[Resources](#) © 2011 Twitter



Feedback



- What types of information would you find useful on the “NCCS User News” feed?
- Is there any information that you prefer not to see in the feed?
- Do you have any questions or other comments about the feed?
- Contact:
Jarrett Cohen
Managing Editor, “NCCS User News”
301-286-2744
jarrett.s.cohen@nasa.gov



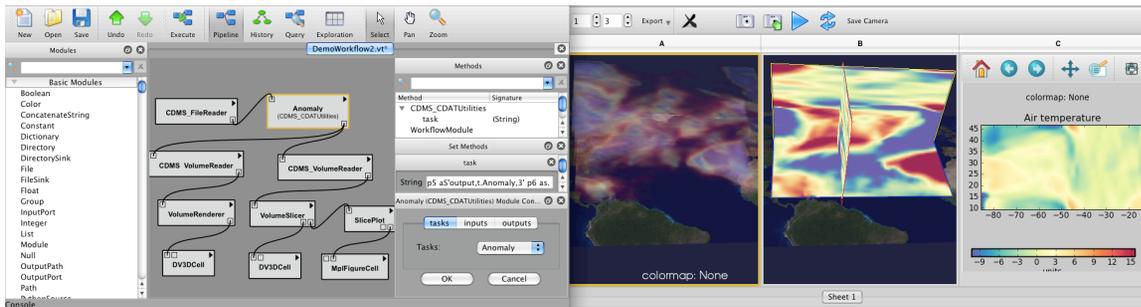
Supporting Slides – Ultra Scale Climate Data Analysis Tools



vtDV3D – Simple Interface, Powerful Analysis, Visualization, and Workflow Tool



NCCS-developed vtDV3D provides a simple GUI interface to advanced visualization workflows and analysis of climate data for scientists using Discover's Dali analysis nodes and the Visualization Wall systems in the NCCS's Data Exploration Theater. (Point of contact: Tom Maxwell/SAIC for NCCS, Thomas.Maxwell@nasa.gov)



Advanced features include:

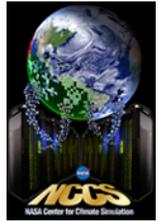
- **vtDV3D provides high-level modules for UV-CDAT, the Earth System Grid visualization and analysis framework from PCMDI.**
- vtDV3D is a package of VisTrails modules, and will be included in the upcoming 2.0 release of VisTrails, an open-source scientific workflow and provenance management package that is supported by DoE's Office of Science and NSF.
- vtDV3D provides simple interactive keyboard and mouse controls for:
 - Dragging slice planes.
 - Resizing and rotating.
 - Adjusting color map scales.
 - Modifying color and opacity of volume render transfer function.
 - Obtaining coordinate & variable values available via mouse picks.
- Volume rendering, 2D slicing, iso-contouring, and streamlining modes.
- Active and passive stereo vision support (via VTK).
- Access to custom VisTrails module GUIs for choosing color map, etc.
- Greatly simplified analysis & visualization workflows.
- Multiple concurrent synchronized visualizations.
- Hyperwall display mode with separate synchronized displays on each hyperwall tile.
- Hyperwall application development on the user's desktop.



Supporting Slides – Discover Utilization

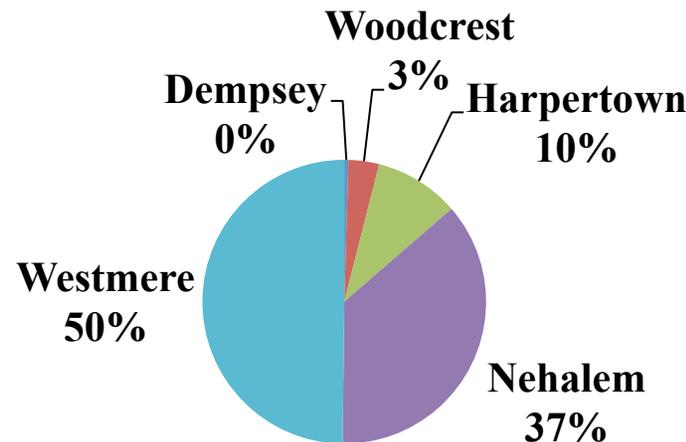


Discover Status



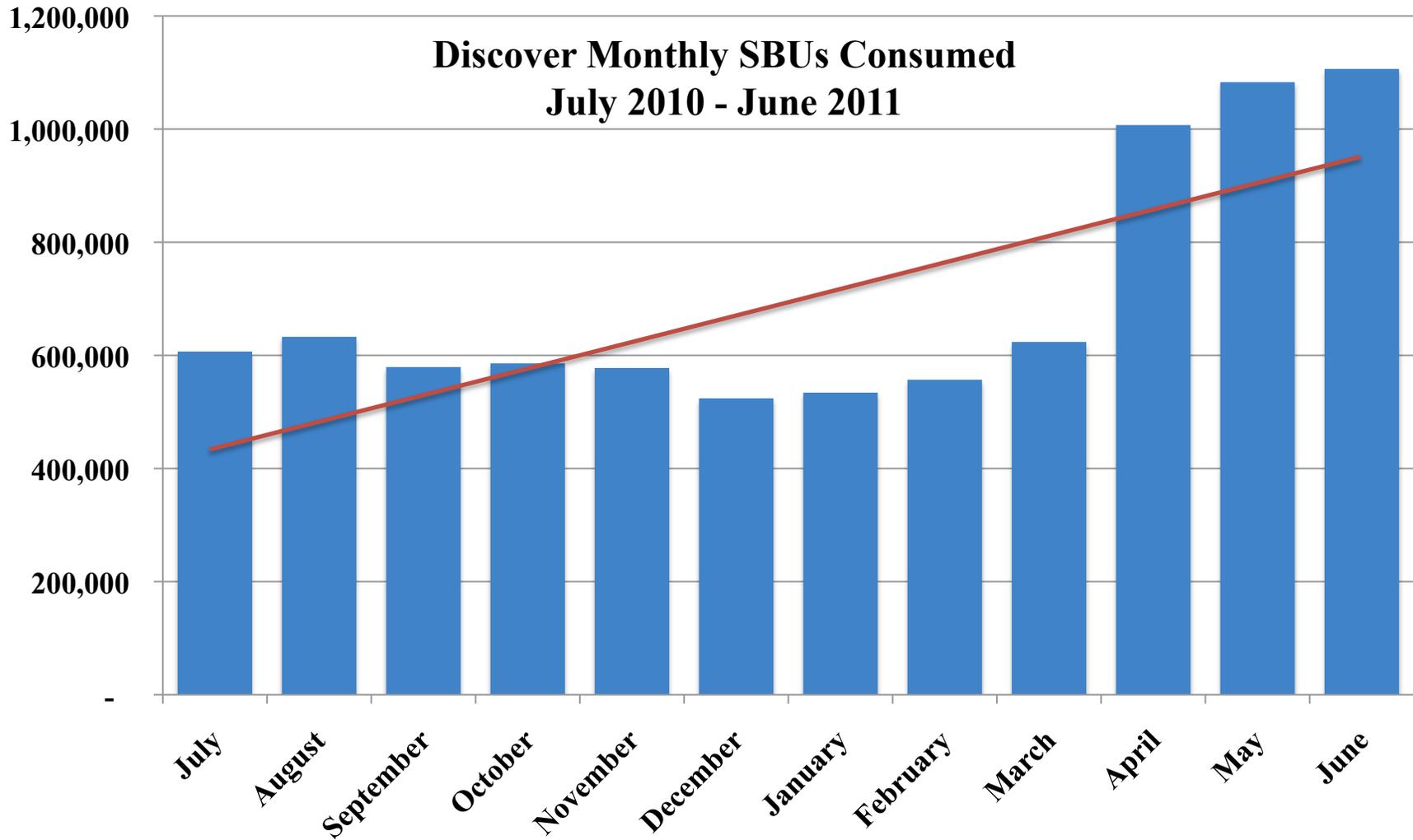
- SCU7 queue configuration changes
- PBS 10, SLES 11 software upgrades
- SCU3, SCU4 hardware upgrades: to Westmere
- Base Unit, SCU1, SCU2 future
- Additional nobackup disk storage
- Improvements on GPFS hangs, nodes run out-of-memory

**Discover Workload Distribution by Processor Type
(Including Dedicated Resources) - June 2011**



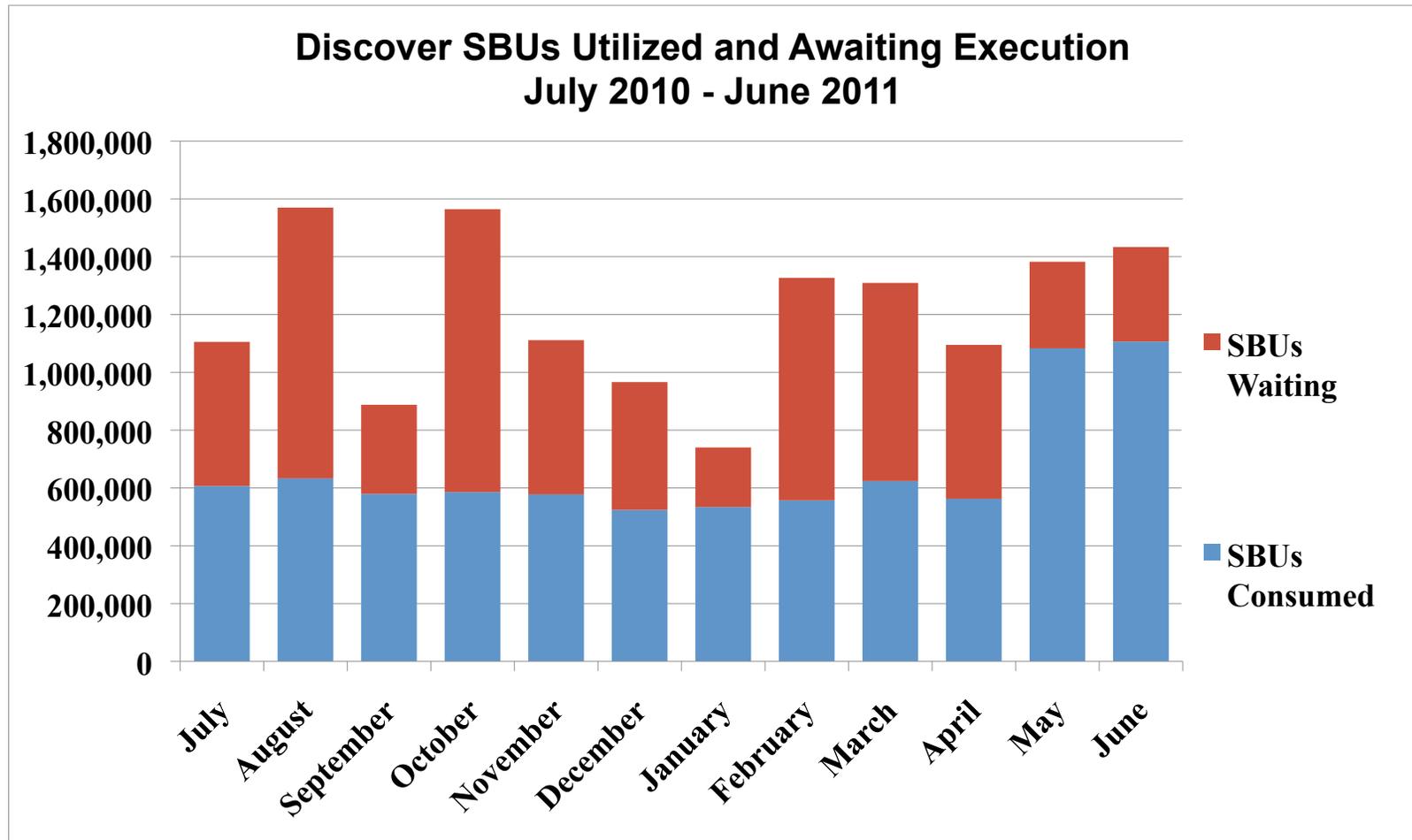


Discover 12-Month SBU Consumption



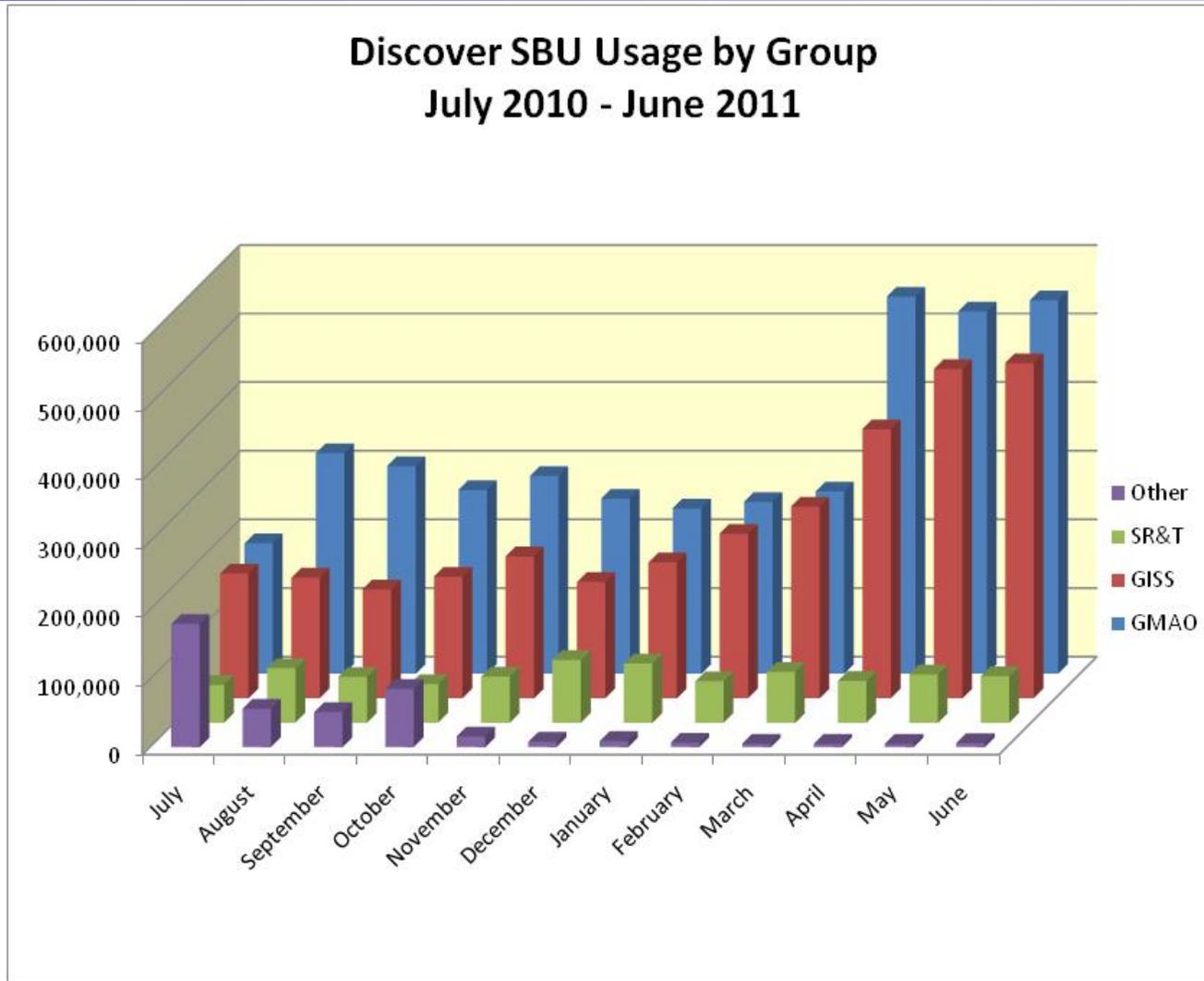


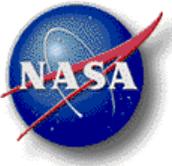
Discover 12-Month SBU Utilization and Wait



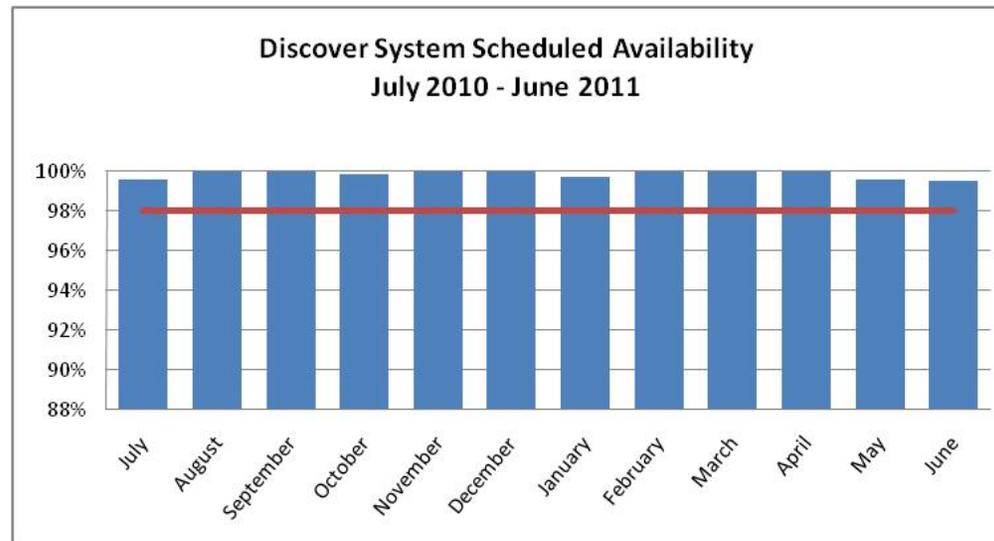
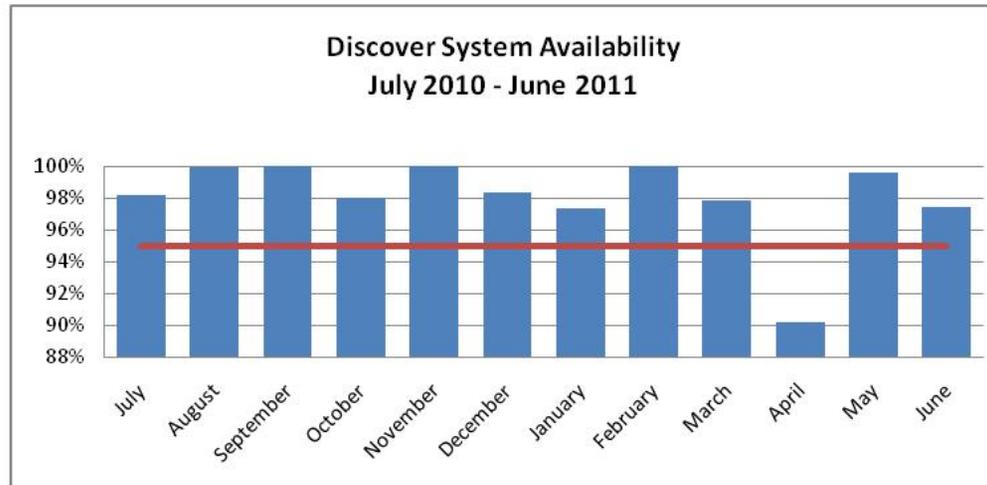


Discover 12-Month SBU Usage by Group





Discover 12-Month Availability





Mass Storage Status



- Improved stability
- SLES 11
- Additional storage
 - Disk
 - Tape

