The TAMU CMS Tier 3 Site at the Brazos Cluster

High Energy Physics and
High Performance Computing at TAMU
Michael Mason

Audience

- This talk is divided into two sections for two different audiences
- People New to Grid Computing
  - If you are new to the ideas of Grid computing the first half of the talk explains some general concepts
  - You will become familiar with how the CMS model works
- Experienced CMS Users
  - The second half of the talk covers all the functionalities of the CMS Grid at our new CMS Tier 3 site
  - We also present information about interacting locally with the Brazos Cluster outside the scope of Grid work

Outline

- Cluster & Grid Computing
  - Computer, Cluster, Grid comparison
  - CMS Tiered Structure
- CMS Grid
  - Access with Certificates
  - Running jobs with CRAB
  - Moving Data with PhEDEx
- TAMU Computing at Brazos
  - Why we are using Brazos
  - Brazos Hardware and Administrative Structure
  - Storage Resources
  - Practical Guides for using the CMS Grid
  - Notes on Local/non-Grid usage
- Questions

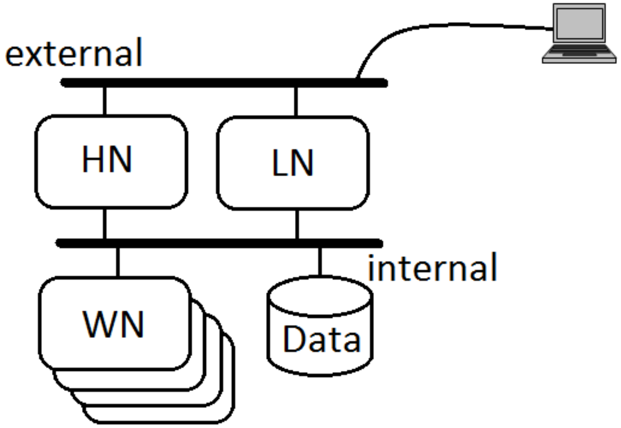
New to CMS Grid

Experienced CMS user

Computer

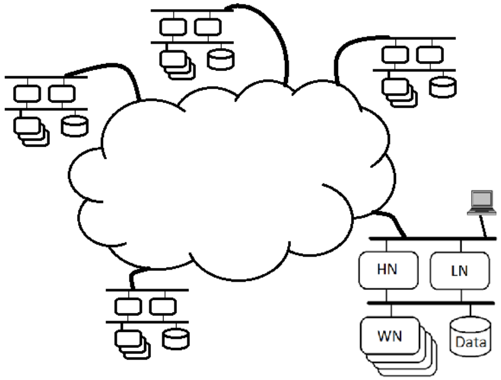
- General Description
  - One box
  - One user
  - One CPU shared by all processes (jobs) on the system
  - Operating System to control the system resources
- User interaction
  - Direct access
- Job submission
  - Entire job is run on single CPU shared by all jobs

Cluster



- General Description
  - Many Computers connected by a local network
  - Many users (~10-100)
  - A few dedicated CPUs for system management and user interaction
  - Many (~100-1000) CPUs dedicated for jobs
  - Local Batch System to control the CPU resources
- User interaction
  - Remote connection to one CPU on the system
- Job submission
  - User submits single job to the system
  - System can make intelligent decision to run job in parallel over multiple CPUs for faster processing

Grid



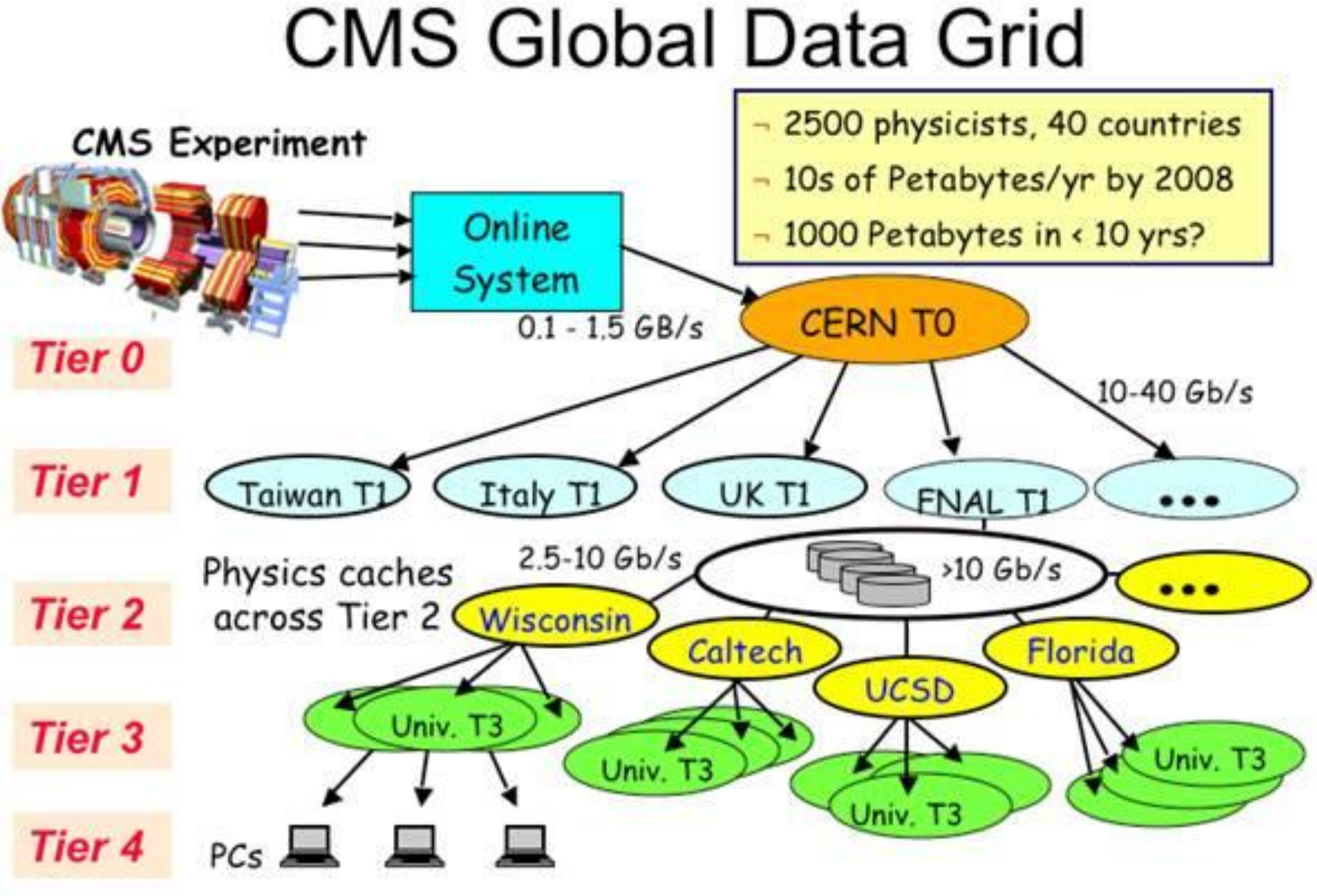
- General Description
  - Many Clusters around the world connected by a wide area network
  - Many users (many thousands)
  - A few dedicated CPUs for grid management per Cluster
  - Each Cluster's CPUs are now shared by all Clusters in the Grid
  - Workload Management System to control the Grid resources
- User interaction
  - User only interacts with their "Cluster", the Grid is a transparent feature
- Job submission
  - User submits single job to the Grid
  - The Grid makes intelligent decisions to run job in parallel over multiple CPUs across one or more Clusters
  - Output data is returned to your local Cluster
  - Datasets can be moved around the Grid as needed

CMS Grid: A Tiered System

- Tier 0
  - CERN: collect raw data from the CMS experiment and packages it into datasets
- Tier 1
  - 7 around the world, ours is FNAL
  - Backup and redistribute datasets to Tier 2 and 3 sites
  - Host primary collaboration Monte Carlo (MC) datasets
- Tier 2
  - Analysis and MC production
  - Must share resources with everyone in CMS
- Tier 3
  - Analysis and MC production (usually smaller scale)
  - No requirement for resource sharing
- Brazos is a Tier 3 site that will give us priority access to a large number of CPUs as well as our own large data storage connected directly to the grid

CMS Tiered Structure

- The Grid shares data between Clusters as well as CPUs



Certificates

- To submit jobs to the Grid you use a certificate for identification instead of having a username/password on every Cluster
- Certificates are issued to users by Certificate Authorities (CA)
  - The CA's typically used to get a certificate in the US are:
    - CERN Certificates (https://ca.cern.ch/ca/)
    - DOE Certificates (http://www.doegrids.org/)
  - Users are verified by the CA, then issued their unique certificate required to access the Grid

CMS Grid Job – CRAB

- All CMS Grid jobs are submitted using a program known as CRAB
  - Cms Remote Analysis Builder
- CRAB takes one job that runs over one dataset and intelligently splits it up into multiple smaller jobs
  - The user configures how the job is split up
- Those many smaller jobs are then submitted by the Grid to one or more Clusters that have the specified dataset
- The output of each job is then returned to the user on their local cluster

PhEDEx – Moving Data within the CMS Grid

- Physics Experiment Data Export (PhEDEx)
  - This is how data is moved around the Grid
- All datasets are registered with the Dataset Bookkeeping System (DBS)
- PhEDEx knows where each dataset is
  - Which site it is at (can be multiple sites)
- Users
  - Can request that datasets be moved to any site
  - Uses a simple web interface
- Data Managers
  - In charge of data storage for their site
  - Can approve or deny any request to move data to their site

The Brazos Cluster at TAMU

- Worker Nodes
  - Full Cluster will have 310 nodes, total of 2,720 cores; each capable of running independent jobs
  - We have Preferred Access to 8 nodes, 256 cores
- Storage
  - 35TB shared by HEPx group (CMS users)
  - ~60TB more shared by all Brazos users
- Stakeholder model
  - Stakeholders buy into the Cluster and are given access to all its resources when needed
  - Stakeholders have preferred access in proportion to their contribution
- Access/Account information later in talk
  - More info at http://brazos.tamu.edu/

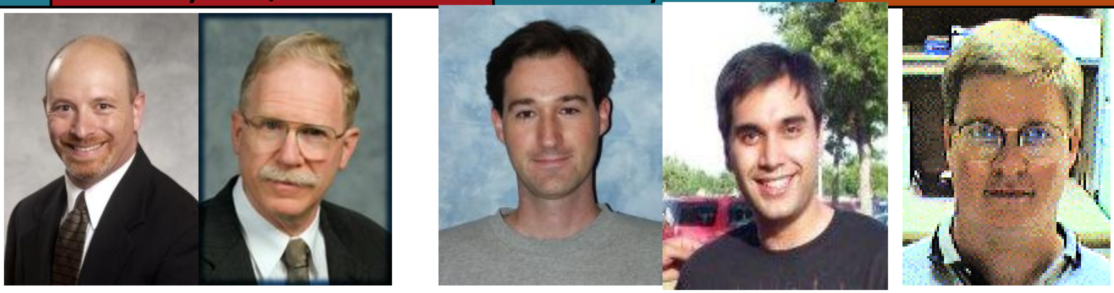
Why use Brazos?

- Our own Cluster
  - Pros
    - Full control over things like batch system, file system, programs
  - Cons
    - Very small cluster
    - Need full time sysadmins
    - Lack experience running cluster and grid based computing

- Brazos Cluster
  - Pros
    - Potential access to 2656 cores and ~100TB of storage if needed
    - Round the clock experienced sysadmins
    - Smaller setup time adding on to an existing cluster
  - Cons
    - Give up control
    - Share resources with other users

Tier 3 Administration

Job:	Principle Investigators	CMS Admin	Grid/OSG Admin	Brazos SysAdmin
Who:	Dave Toback	Mike Mason	Steve Johnson	Many people
	Guy Almes	Vaikunth Thukral		brazos-help@listserv.tamu.edu
Respon-sibility:	Project Leadership Future Planning Funding	CMSSW CRAB PhEDEx	Compute Element Storage Element	Security Software/TORQUE Maintenance
From:	Physics/Brazos	Physics	Brazos	



More info on: http://collider.physics.tamu.edu/tier3/

- A mix of Brazos and Physics people
- Talk to CMS Admins when:
  - You have a problem with any CMS software or need a new version
- Talk to the Brazos Admins when:
  - You have questions about the Cluster or want to run local jobs
  - You want a particular type of non-CMS software installed

Disk Space and Directory Structure on Brazos

- Home directory
  - /home/USERNAME
  - 10GB backed up
- HEPx Data (coming soon)
  - Shared by HEPx group only
  - 35TB not backed up
  - All PhEDEx transferred datasets
    - /data/hepxadmin/store
  - CRAB data copied back to our site T3_US_TAMU
    - /data/hepxadmin/store/user/USERNAME
  - User storage for unpublished MC/testing/analysis
    - /data/USERNAME
    - Mixed with the shared 60TB from Brazos

Disk Space Usage on Brazos

- We expect the bulk of our storage needs will be in datasets registered in the DBS
- Propose that PhEDEx datasets be assigned to Groups for internal bookkeeping purposes
  - Example: SUSY, Tau, and SLHC could all be groups that need to share datasets
- Will periodically monitor datasets as well as CRAB and User spaces
- Any Group datasets should be deleted when no longer needed
- At the moment there is no strict storage limits but quotas may need to be implemented if we start to run low on storage space

How-To Guides for Users

There are How-To guides covering information in the rest of the talk that can be found at:

- For all TAMU Tier 3 information
  - http://collider.physics.tamu.edu/tier3/
- For specific CMS information
  - https://twiki.cern.ch/twiki/bin/view/CMS/WorkBook
- For information about Brazos
  - http://brazos.tamu.edu/docs.html

Changing Gears

- This is the dividing line in the talk where we switch from general Grid information to specific CMS information for the new Tier 3 site
- At this point we are assuming you are familiar with CMS software and general CRAB and PhEDEx usage

CRAB: Info for Users

- For more info see the CRAB guide
  - http://collider.physics.tamu.edu/tier3/CRAB-User_Guide.pdf
- Cms Remote Analysis Builder
  - Submit jobs to the Grid
  - Intelligently splits your one big job into many small jobs and sends them to sites with the dataset you requested
- Need your Certificate
  - Add your usercert.pem and userkey.pem to ~/.globus
- Setup your environment in the right order
  - source /home/hepxadmin/gLite/gLite-UI/etc/profile.d/grid_env.sh
  - cmsenv
  - source /home/hepxadmin/crab/current/crab.sh
- Run CRAB as normal
  - https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookRunningGrid
- Use T3_US_TAMU as your return SE
- You MUST whitelist T3_US_TAMU in your crab.cfg file for any job to run at Brazos
- Brazos only accepts Grid jobs from gLite (not glidein)

PhEDEx: Info for Users

- More info at
  - http://collider.physics.tamu.edu/tier3/ PhEDEx-User_Guide.pdf
- All datasets are registered with DBS (Dataset Bookkeeping System)
- PhEDEx knows where each dataset is
  - i.e. Which site it is at (can be multiple sites)
- PhEDEx is not real time!
  - Must have a link between sites to transfer data
  - It may take a few days to get a link setup, after which it is permanent
  - Current links T1: FNAL, CERN; T2: Caltech, Florida, MIT, Nebraska, Purdue, UCSD, Wisconsin; T3: Vanderbilt
- Plan ahead!
  - Don't expect an obscure 2TB dataset to be here in an hour
  - Approximate transfer rate is 1TB/6 hours (varies per site)
- Using PhEDEx
  - Simple Web Interface
  - From DBS search (https://cmsweb.cern.ch/dbs_discovery/)
  - Directly (http://cmsweb.cern.ch/phedex)

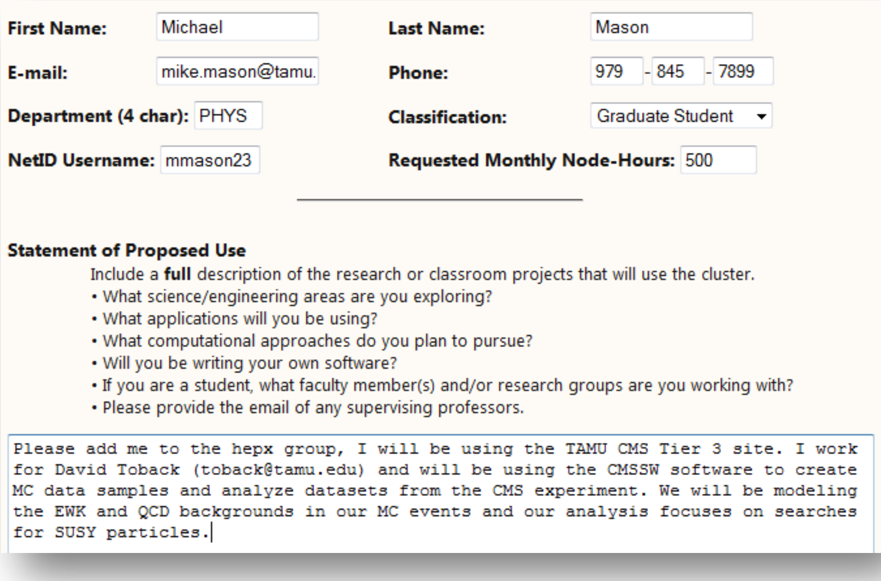
Local Brazos Usage: Non-Grid

We next describe some non-Grid usage of Brazos including:

- Getting an Account on Brazos
- Setting up CMSSW on Brazos
- Local batch queue: Torque
- Local interactive jobs

Brazos Access

- Creating an Account
  - Follow Brazos-Account Guide
    - http://collider.physics.tamu.edu/tier3/Brazos_Account_Guide.pdf
  - http://brazos.tamu.edu/access.html
  - Read and accept their policies, login using NetID
  - Fill in your TAMU info, leave defaults
  - Statement of Proposed Use
    - Must be added to hepx group
    - Include advisor and their contact info
    - Using CMS Tier 3 facilities
    - 2 to 3 sentences about your research
- Your account will be created in about 1 day
- Accessing the Cluster
  - ssh -X username@brazos.tamu.edu
  - ssh -X username@hurr.tamu.edu



First Name:	Michael	**Last Name:**	Mason
E-mail:	mike.mason@tamu.	**Phone:**	979 - 845 - 7899
Department (4 char):	PHYS	**Classification:**	Graduate Student ▾
NetID Username:	mmason23	**Requested Monthly Node-Hours:**	500

Statement of Proposed Use

Include a **full** description of the research or classroom projects that will use the cluster.
- What science/engineering areas are you exploring?
- What applications will you be using?
- What computational approaches do you plan to pursue?
- Will you be writing your own software?
- If you are a student, what faculty member(s) and/or research groups are you working with?
- Please provide the email of any supervising professors.

```
Please add me to the hepx group, I will be using the TAMU CMS Tier 3 site. I work
for David Toback (toback@tamu.edu) and will be using the CMSSW software to create
MC data samples and analyze datasets from the CMS experiment. We will be modeling
the EWK and QCD backgrounds in our MC events and our analysis focuses on searches
for SUSY particles.
```

CMSSW on Brazos

- For more information see the CMSSW-User Installation Guide
  - http://collider.physics.tamu.edu/tier3/CMSSW-User_Installation_Guide.pdf
- Setting up your environment
  - Edit .bash_profile
  - export VO_CMS_SW_DIR=/home/hepxadmin/cmssw
  - export SCRAM_ARCH=slc5_ia32_gcc434
  - source $VO_CMS_SW_DIR/cmsset_default.sh
- For CMSSW versions < 3_5_0
  - export SCRAM_ARCH=slc4_ia32_gcc345
- Now use SCRAM as normal to get and build CMSSW
  - Source Configuration, Release, And Management tool
- scram list CMSSW
  - See CMSSW versions for your SCRAM_ARCH
- scram project CMSSW <CMSSW_Version>
  - To install a version
- Contact CMS Admins if a new version is needed

Running locally on Brazos – TORQUE

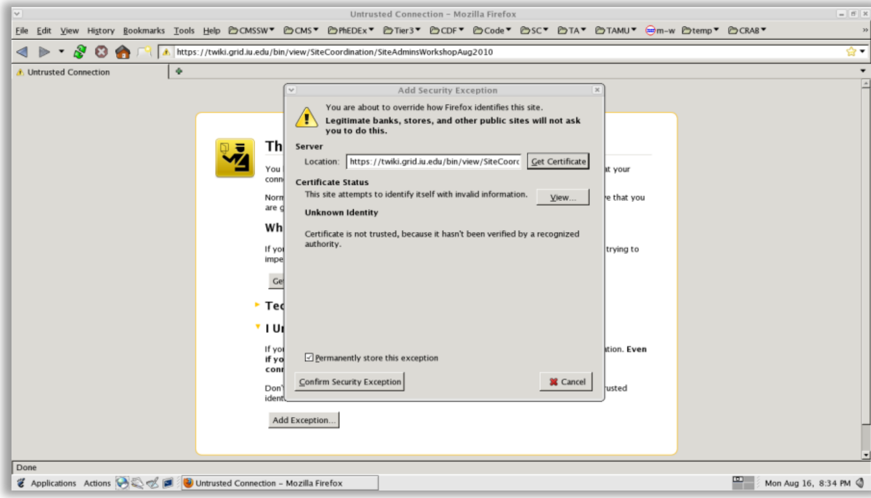
- Local Batch System used by Brazos
  - TORQUE uses Portable Batch System (PBS), you may be familiar with Condor
  - This is a Cluster activity, not a Grid activity, CMS Admins can help but are not experts on this
- Used to submit jobs to run locally on the Compute Nodes
- See the Brazos guide for details
  - http://collider.physics.tamu.edu/tier3/CMSSW-Brazos_Guide.pdf

Local Interactive Jobs

- Get interactive access to a Worker Node
  - Useful for quick CMSSW testing
  - CMSSW is too intensive to run on a Login Node
  - REMEMBER TO LOG OFF
    - No other jobs can run on that node while you are logged in
- qsub –I –X –V –d $PWD –q hepx
  - –I = interactive job
  - –X (optional) = Enables X11 forwarding
  - –V (optional) = export all environment variables
  - –d path (optional) = working path
  - –q queue = we own the hepx queue
- Root is much less intensive
  - ROOT analysis can be run directly on the Hurr Login Node
  - Don’t use the Brazos Login Node, too many people work there
  - Again, CMS admins can help but are not experts in this

Helpful Tip: Adding Trusted CAs to your Browser

- The CERN and DOE Certificate Authority (CA) are not trusted by default but are used on many CMS web pages
- Don’t add exceptions ⇒
- Add CAs to avoid warning messages on web sites
- To recognize CAs
  - https://www.tacar.org/repos/
  - Install all CERN and DOE certificates


Untrusted Connection – Mozilla Firefox
File Edit View History Bookmarks Tools Help CMSSW CMS PhEDEx Tier3 CDF Code SC TA TAMU m-w temp CRAB
https://twiki.grid.iu.edu/bin/view/SiteCoordination/SiteAdminsWorkshopAug2010
Untrusted Connection
Add Exception...
Done
Applications Actions
Untrusted Connection – Mozilla Firefox
Mon Aug 16, 8:34 PM

Add Security Exception

You are about to override how Firefox identifies this site.

Legitimate banks, stores, and other public sites will not ask you to do this.

Server

Location: https://twiki.grid.iu.edu/bin/view/SiteCoorc

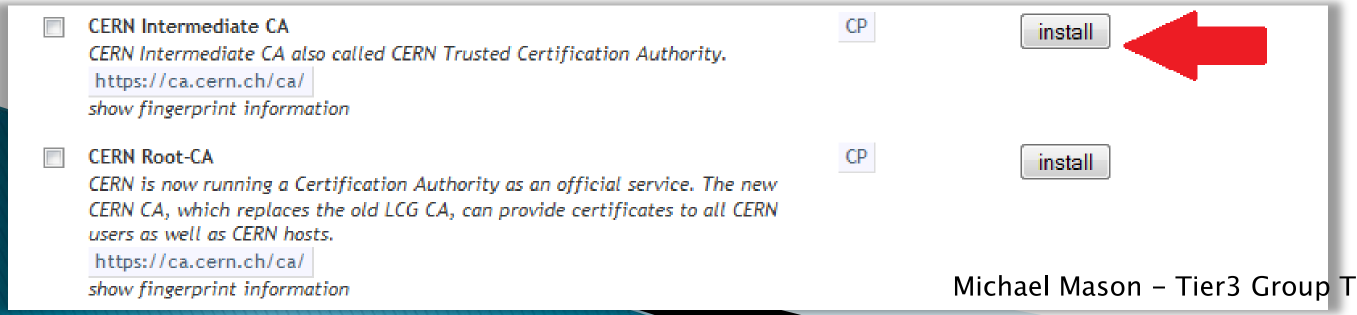
Certificate Status

This site attempts to identify itself with invalid information.

Unknown Identity

Certificate is not trusted, because it hasn't been verified by a recognized authority.

☑ Permanently store this exception



☐ CERN Intermediate CA

CERN Intermediate CA also called CERN Trusted Certification Authority.

https://ca.cern.ch/ca/

show fingerprint information

☐ CERN Root-CA

CERN is now running a Certification Authority as an official service. The new CERN CA, which replaces the old LCG CA, can provide certificates to all CERN users as well as CERN hosts.

https://ca.cern.ch/ca/

show fingerprint information

Summary

- The TAMU CMS Tier 3 is now operational and running on the Brazos Cluster
- We have access to a large number of CPUs and disk space for our jobs
- We have the ability to use CRAB to intelligently split up our jobs and run them around the Grid
- We can use PhEDEx to manually move datasets around the Grid and here to TAMU
- Using the Brazos Cluster offloads much of the administration work
- By joining the Brazos Cluster we have more disk space, processing power and support than we would have on our own
- Please use and enjoy this powerful new tool

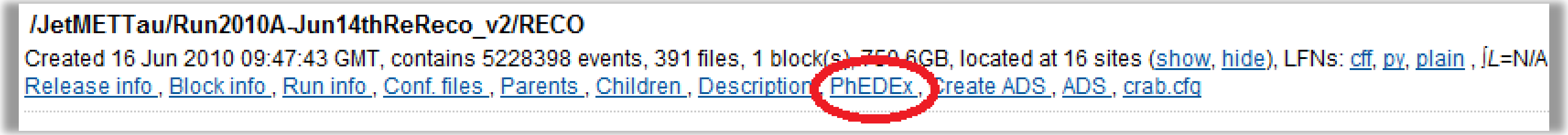
Questions – For You

- Are there ways you typically run that we are not well set up for?
- CMSSW versions we have installed
  - 3.1.6
    - SCRAM_ARCH=slc4_ia32_gcc345
  - 3.6.2, 3.6.3, 3.7.0(patch3 and 4), 3.8.5
    - SCRAM_ARCH=slc5_ia32_gcc434
  - Need more?
  - We will soon switch to an automated CMSSW updater that will include all (non-deprecated) versions
- We have PhEDEx links to
  - T1: FNAL, CERN
  - T2: Caltech, Florida, MIT, Nebraska, Purdue, UCSD, Wisconsin
  - T3: Vanderbilt
  - Where else?
- Comments or concerns about our plan for keeping track of the data space usage and deleting old datasets?

Backup

Using PhEDEx

- Simple Web interface (Login with Certificate)
  - Search for data at DBS (https://cmsweb.cern.ch/dbs_discovery/)

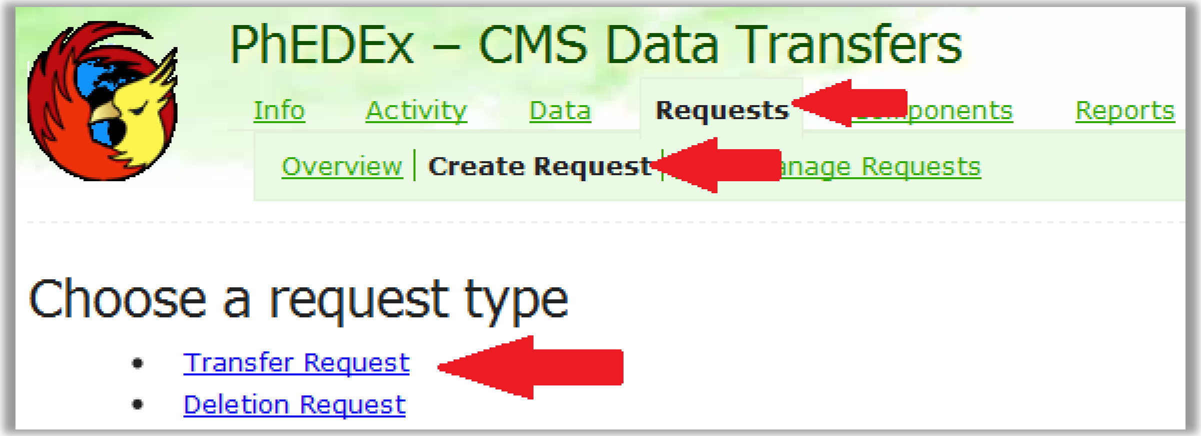


/JetMETTau/Run2010A-Jun14thReReco_v2/RECO

Created 16 Jun 2010 09:47:43 GMT, contains 5228398 events, 391 files, 1 block(s), 750.6GB, located at 16 sites (show, hide), LFNs: cff, py, plain , ∫L=N/A

Release info , Block info , Run info , Conf. files , Parents , Children , Description , PhEDEx , Create ADS , ADS , crab.cfg

  - Or use PhEDEx site (http://cmsweb.cern.ch/phedex)



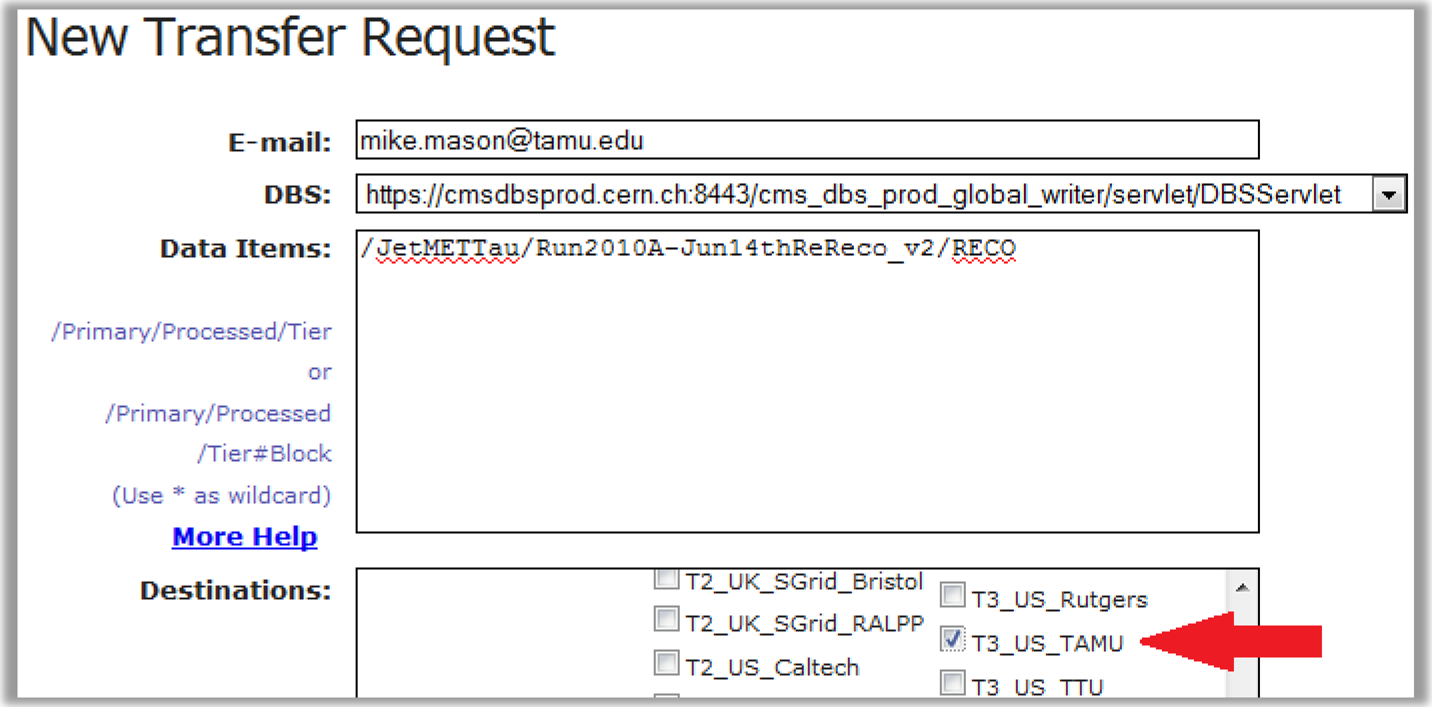
PhEDEx – CMS Data Transfers

Choose a request type

- Transfer Request
- Deletion Request

PhEDEx Transfer Request

- The DBS or you will fill in the dataset
- Select T3_US_TAMU as the destination



New Transfer Request

E-mail: mike.mason@tamu.edu

DBS: https://cmsdbsprod.cern.ch:8443/cms_dbs_prod_global_writer/servlet/DBSServlet

Data Items: /JetMETTau/Run2010A-Jun14thReReco_v2/RECO

/Primary/Processed/Tier or /Primary/Processed/Tier#Block (Use * as wildcard)

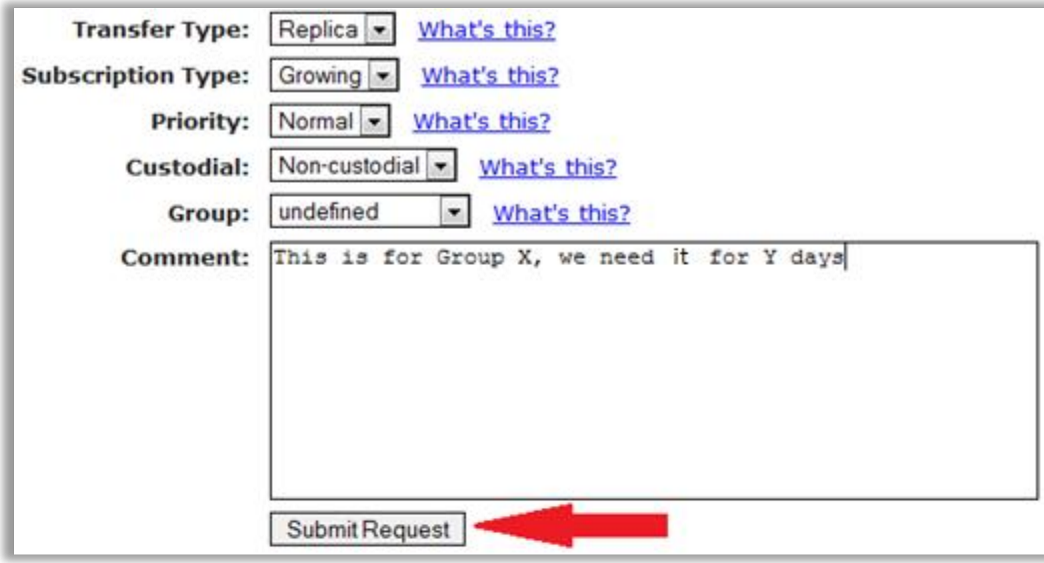
More Help

Destinations:

□ T2_UK_SGrid_Bristol
□ T2_UK_SGrid_RALPP
□ T2_US_Caltech
□ T3_US_Rutgers
☑ T3_US_TAMU
□ T3_US_TTU

PhEDEx Transfer continued

- Keep the defaults
- Include a comment
  - For which group
  - Keep for how long
- Requests must be approved by CMS Admins
- You’ll get an email when your request is (dis)approved



Transfer Type: Replica

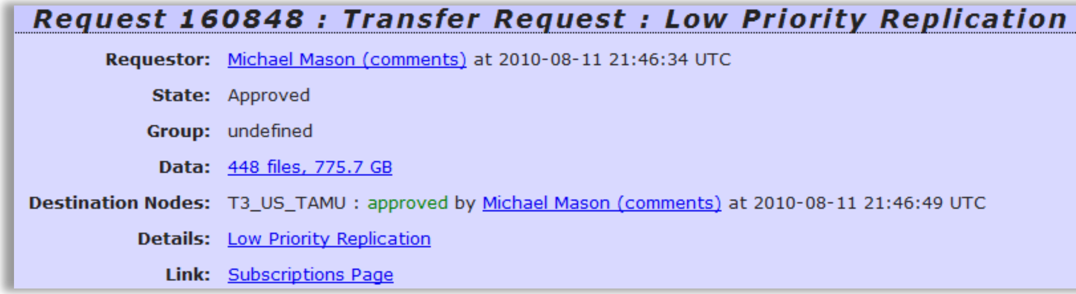
Subscription Type: Growing

Priority: Normal

Custodial: Non-custodial

Group: undefined

Comment: This is for Group X, we need it for Y days



Request 160848 : Transfer Request : Low Priority Replication

Requestor: Michael Mason (comments) at 2010-08-11 21:46:34 UTC

State: Approved

Group: undefined

Data: 448 files, 775.7 GB

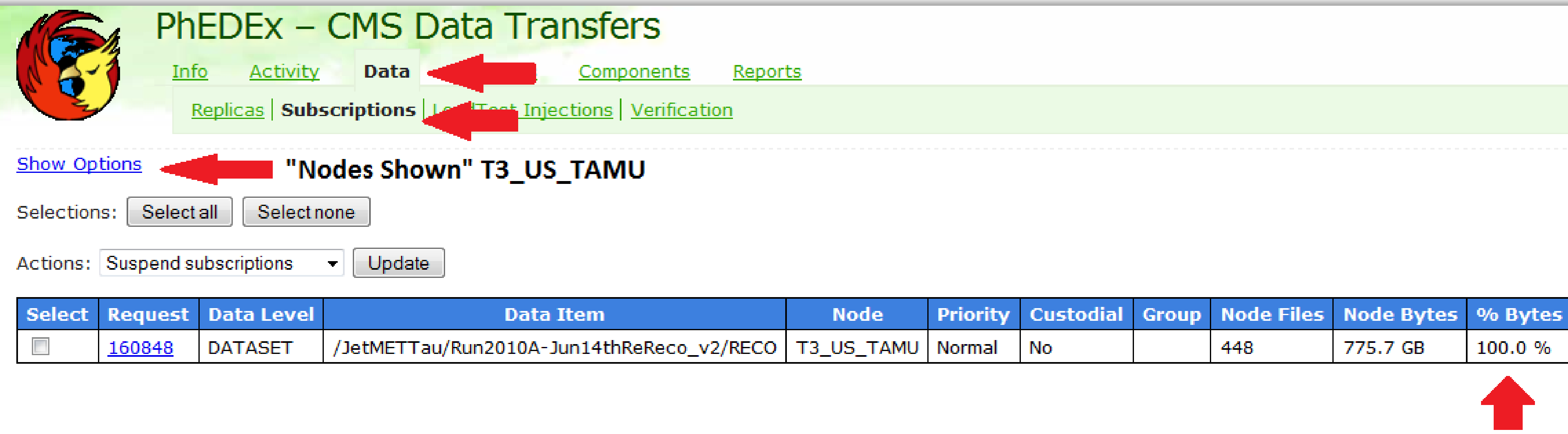
Destination Nodes: T3_US_TAMU : approved by Michael Mason (comments) at 2010-08-11 21:46:49 UTC

Details: Low Priority Replication

Link: Subscriptions Page

PhEDEx Transfer Status

- http://cmsweb.cern.ch/phedex
- View what datasets we have and the progress of transfers



"Nodes Shown" T3_US_TAMU

Selections:

Actions: Suspend subscriptions

Select	Request	Data Level	Data Item	Node	Priority	Custodial	Group	Node Files	Node Bytes	% Bytes
□	160848	DATASET	/JetMETTau/Run2010A-Jun14thReReco_v2/RECO	T3_US_TAMU	Normal	No		448	775.7 GB	100.0 %

- For more information view the PhEDEx guide
  - http://collider.physics.tamu.edu/tier3/