

# Introduction to PhEDEx for Tier-3s

Paul Rossman

- ➡ What is PhEDEx?
- ➡ Current PhEDEx developers
  
- ➡ PhEDEx Instances
- ➡ PhEDEx Agents
  
- ➡ Installing PhEDEx for the first time
- ➡ How to monitor PhEDEx
  
- ➡ PhEDEx support resources
- ➡ Additional resources and links
- ➡ Questions?

# What is PhEDEx?



- ➡ PhEDEx is short for **Physics Experiment Data Export**
- ➡ PhEDEx was developed to manage the movement of data within CMS
- ➡ Provides a centralized system for making global data movement decisions
- ➡ Also provides site and data managers and users a realtime view of the global CMS data transfer state
- ➡ Lastly, PhEDEx automates for CMS many of low level tasks typically found in HEP experiments such as
  - large-scale data replication
  - tape migration
  - consistency
- ➡ In production since mid-2004

# What is PhEDEx? (continued)



- ➡ PhEDEx is “composed of a series of autonomous, robust, persistent, stateless process” (\*) or in PhEDEx terms agents
- ➡ These agents “share information about replica and transfer state through a blackboard architecture” (\*)
- ➡ The blackboard is comprised of a database and other agents at CERN
- ➡ These agents when combined form an instance
- ➡ There are currently 3 instances in use today, Production, Debug and Development
- ➡ Instances are independent and do not share information between each other

(\*) The CMS PhEDEx System: a Novel Approach to Robust Grid Data Distribution

# Current PhEDEx developers



- ➡ Over the years, many have contributed to the development of PhEDEx
- ➡ Current development work is being done by
  - Ricky Egeland (CERN)
  - Tony Wildish (CERN)
  - Chih-Hao Huang (FNAL)

- ➡ There are three PhEDEx instances
  - Production
  - Debug
  - Development
- ➡ Each instance is comprised of several, often similar or duplicate named agents
- ➡ The Production and Debug instances are critical for all CMS sites

# PhEDEx Instances (continued)



- ➡ Production instance (Prod)
  - All CMS experiment data is transferred using this instance
  - For a T3, should try and be operational 24x7
- ➡ Debug instance (Debug)
  - Critical instance used for site commissioning and readiness
  - Tier-1 and Tier-2 sites transfer links can be decommissioned if transfer problems in this instance are not resolved
  - The primary focus of this instance was until recently to *debug* transfer agents and links
  - For a T3, should try and be operational 24x7
- ➡ Development instance (Dev)
  - Was historically used during testing and validation of PhEDEx software releases
  - Plans are being developed to re-task this instance to debugging

- ➔ Each instance is comprised of several agents, for example
  - FileDownload
  - FileExport
  - FileStager
  - FileRemove
  - FileMSSMigrate
  - BlockDownloadVerify
- ➔ Not all sites require all agents
  - Sites without an MSS do not require the FileMSSMigrate agent
- ➔ Agent status for the various instances are available at
  - <http://cmsweb.cern.ch/phedex/prod/Components::Status>
  - <http://cmsweb.cern.ch/phedex/debug/Components::Status>
  - <http://cmsweb.cern.ch/phedex/test/Components::Status>



# PhEDEx Agents (continued)



- ➡ For Tier-3s, the minimum agent configuration in each instance should consist of these 4 agents
  - FileDownload
  - FileExport
  - FileRemove
  - BlockDownloadVerify

# Installing PhEDEx for the first time



- ➡ The instructions on the following pages are paraphrased from
  - <https://twiki.cern.ch/twiki/bin/view/CMS/PhedexAdminDocsNewSite>
- ➡ Please read all the documentation prior to starting the install process
- ➡ Minimum requirements
  - One moderately equipped SL4 Linux machine for PhEDEx
  - Configured SE (presumably Bestman)
  - Some form of storage, with enough disk space to store your data



# Initial PhEDEx setup - Personal Affairs



1. Become CERN user and CMS member
  - <http://ph-dep-usersoffice.web.cern.ch/ph-dep-UsersOffice/>
2. Acquire a CERN computer account
  - <http://it-div.web.cern.ch/it-div/documents/ComputerUsage/CompAccountRegistrationForm-English.pdf>
3. Gain access rights in CMS CVS
  - [https://savannah.cern.ch/support/?func=additem&group=cmscompinfrasup&assigned\\_to=3534&category\\_id=107](https://savannah.cern.ch/support/?func=additem&group=cmscompinfrasup&assigned_to=3534&category_id=107)
4. Acquire a certificate from DOEGrids cert
  - <http://www.doegrids.org/>
5. Register with LCG
  - <http://lcg.web.cern.ch/LCG/registration.htm>
6. Register your DOEGrid cert with LCG User Registration
  - <https://ca.cern.ch/ca/>
7. Get a HyperNews account
  - <https://twiki.cern.ch/twiki/bin/view/CMS/PhedexAdminDocHyperNewsForums>
8. Register with SiteDB
  - <https://cmsweb.cern.ch/sitedb/sitedb/>

# Initial PhEDEx setup - Site Affairs



- ➔ Open a Savannah ticket at <https://savannah.cern.ch/support/?func=additem&group=cmscompinfrasup> with the following details
- Site name (e.g. Pisa)
  - CMS name (e.g. T2\_IT\_Pisa) Please keep the length of the CMS name to 20 characters or less
  - SAM name. This should be the same thing as the GlueSiteName attribute published in BDII. For EGEE sites, this corresponds to the site name in GOCDB (e.g. INFN-PISA for T2\_IT\_Pisa)
  - City/Country
  - Site tier (e.g. Tier 2)
  - Storage Element host (e.g. cmsdcache.pi.infn.it)
  - Storage Element kind (e.g. Buffer, MSS, Disk)
  - Storage Element technology (e.g. dCache, Castor, DPM, Disk, Other)
  - Computing Element host(s) (e.g. gridce2.pi.infn.it gridce.pi.infn.it )
  - Associated Tier I (e.g. INFN)
  - Grid type (e.g. LCG, OSG)
  - Data Manager name, e-mail, phone, certificate DN
  - PhEDEx Contact name, e-mail, phone, certificate DN
  - Site Admin name, e-mail, phone, certificate DN
  - Site Executive name, e-mail, phone, certificate DN



# Now it's time to actually install PhEDEx



- ➡ Detailed instructions are available from
  - <https://twiki.cern.ch/twiki/bin/view/CMS/PhedexAdminDocsInstallation>
- ➡ The process is basically as follows
  - Create a user on your PhEDEx node 'phedex' or 'cmsprod' etc
  - Create a directory inside your new users home directory (for instance ~/phedex)
  - Cut and paste the test from the next slide following (some tweaking may be necessary)
- ➡ By this point, you should have received 3 PhEDEx credentials via email for connecting to the PhEDEx database from the steps on slide 11
- ➡ Instructions on credential installation are included in the email



# Now it's time to actually install PhEDEx



**Note:** For 64 bit installs, set `myarch=slc4_amd64_gcc345`

```
mkdir ~/phedex
cd ~/phedex
mkdir -p state logs sw gridcert
chmod 700 gridcert
export sw=$PWD/sw
version=3_1_2
myarch=slc4_ia32_gcc345
wget -O $sw/bootstrap.sh http://cmsrep.cern.ch/cmssw/cms/bootstrap.sh
sh -x $sw/bootstrap.sh setup -path $sw -arch $myarch 2>&1tee $sw/bootstrap_$myarch.log
source $sw/$myarch/external/apt/*/etc/profile.d/init.sh
apt-get update
apt-get install cms+PHEDEX+PHEDEX_$version
rm -f PHEDEX; ln -s $sw/$myarch/cms/PHEDEX/PHEDEX_$version PHEDEX
```

# Configuring PhEDEx



- ➡ Assuming all Personal and Site affairs are complete...
- ➡ First create your siteconf directory
  - `~/SITECONF/T3_US_Yoursite/PhEDEx`
- ➡ Then copy the configuration templates files to your new siteconf directory from
  - `~/phedex/PHEDEX/Custom/Template/`
- ➡ Or my preference, review other T3 siteconfs on the PhEDEx CVS repo and build your own
  - <http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/COMP/SITECONF/>
- ➡ Copy what you need and customize to your site
- ➡ Check in your siteconf files to the PhEDEx CVS repo
  - This is important not only as a backup, but to allow other PhEDEx admins the opportunity to review your configurations if needed for debugging assistance



# PhEDEx agents download agents



- ➡ Download and install Fermilab SRM client tools
  - <https://srm.fnal.gov/twiki/bin/view>
- ➡ Some documentation on srmcp available from
  - <https://twiki.grid.iu.edu/bin/view/Documentation/StorageSrmcpUsing>
- ➡ For Tier-3s you will probably want to use Brian Bockelman's srmcp wrapper script
  - <http://t2.unl.edu/documentation/srmcp-t3>



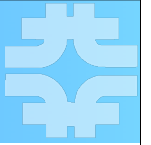
- ➡ Paraphrased from
  - [https://twiki.cern.ch/twiki/bin/view/CMS/PhedexAdminDocsInstallation#Certificate\\_Management](https://twiki.cern.ch/twiki/bin/view/CMS/PhedexAdminDocsInstallation#Certificate_Management)
- ➡ To run the transfer agents, you need to have a grid certificate and be registered to the CMS VO (you did this a few slides back)
  - Note that LCG policy states that all file transfers must be done with personal certificates, not service certificates
- ➡ On a weekly basis, renew your proxy with the following two commands
  - `voms-proxy-init -valid 190:00 -voms cms`
  - `myproxy-init -d -c 720 -s myproxy-fts.cern.ch`

# Starting PhEDEx



- ➡ We first need to test your Debug link for commissioning
- ➡ The Master script for your Debug instance
  - `~/phedex/PHEDEX/Utilities/Master -config ~/SITECONF/T3_US_Yoursite/PhEDEx/Config.Debug start`
- ➡ At this point, please update your original new site Savannah request asking that link commissioning be started
- ➡ Once link commissioning is complete, the Production instance can be started
  - `~/phedex/PHEDEX/Utilities/Master -config ~/SITECONF/T3_US_Yoursite/PhEDEx/Config.Prod start`
- ➡ The link commissioning process needs to be repeated for each new link you request

# How to monitor PhEDEx



- ➡ PhEDEx keeps detailed logs for all agents and transfers in directories specific to each instance
  - `~/phedex/INSTANCE_NAME`
- ➡ Some PhEDEx web pages for monitoring
  - <http://cmsweb.cern.ch/phedex/prod/Activity::Rate> (be sure to click the *Include links with nothing but errors* box)
  - <http://cmsweb.cern.ch/phedex/prod/Activity::QualityPlots>
  - <http://cmsweb.cern.ch/phedex/prod/Components::Status>
- ➡ Make sure you have selected the instance you want to monitor



# PhEDEx support resources



- ➡ PhEDEx documentation and other documents can be found at
  - <https://twiki.cern.ch/twiki/bin/view/CMS/PhEDEx>
- ➡ Questions regarding PhEDEx can be posted to one of these HyperNews forums
  - [hn-cms-phedex@cern.ch](mailto:hn-cms-phedex@cern.ch)
  - [hn-cms-osg-tier3@cern.ch](mailto:hn-cms-osg-tier3@cern.ch)
- ➡ Other transfer or site related questions can be posted to these
  - [hn-cms-dataops@cern.ch](mailto:hn-cms-dataops@cern.ch)
  - [hn-cms-facilitiesOps@cern.ch](mailto:hn-cms-facilitiesOps@cern.ch)
- ➡ Transfer problems should be posted to Savannah
  - <https://savannah.cern.ch/projects/cmscompinfrasup/>
- ➡ Bug reports and feature requests should be posted to Savannah
  - <https://savannah.cern.ch/projects/phedex/>
- ➡ Please make sure you have both a HyperNews and Savannah account before posting (please don't post anonymously to Savannah)

# Additional resources and links

- ➡ Very good Tier-3 specific links
  - <http://hep-t3.physics.umd.edu/HowToForAdmins.html>
  - [http://hep.phys.utk.edu/wiki/PhEDEx\\_Installation](http://hep.phys.utk.edu/wiki/PhEDEx_Installation)
  
- ➡ Example configurations
  - [http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/COMP/SITECONF/T3\\_US\\_UMD/PhEDEx/](http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/COMP/SITECONF/T3_US_UMD/PhEDEx/)
  - [http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/COMP/SITECONF/T1\\_US\\_FNAL/PhEDEx/](http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/COMP/SITECONF/T1_US_FNAL/PhEDEx/) (look at the T3 configs)
  
- ➡ <https://cmsweb.cern.ch/phedex/datasvc/doc>
- ➡ <https://cmsweb.cern.ch/overview/session>
- ➡ <http://t2.unl.edu/phedex/xml>



# Questions?

