FermiGrid School

FermiGrid 201
Scripting and running Grid Jobs

Course Outline

- Introduction and Conventions
- Authorization and Credentials (Lab 1)
- Grid Submission: globus-job-run and globus-url-copy (Lab 2)
- Grid Submission: condor_submit, condor_q (Lab 3)
- Submitting jobs to FermiGrid and Open Science Grid (Lab 4)
- Advanced Topics (Lab 5)
  - Globus RSL
  - DAGman
  - DOEGrids certificates
  - Problem Diagnosis
  - Globus Web Services

Introduction

- Start with simple examples that work
  - Go back to show extra tricks often used in production.
- This course will cover examples of submitting jobs from client machine fnpcsrv1 to compute resource fngp-osg.
- By the end of this course you should be able to:
  - Submit a simple job to the grid
  - Submit a complex job to the grid
  - Transfer files to the grid resource
- The examples used here should be good on any Open Science Grid site
  - Introductory material on using other sites in the Appendix.
- You could install your own client on your own machine—Covered in FermiGrid 401.
- Ask lots of questions—we will fill them in and add them to the course in future..
- There is much more material in these slides and handouts than what we will cover in 3 hours, but we added it based on questions that were asked last time.

Conventions

- In each section we will cover:
  - What we want to do in plain English
  - Technical tools and concepts that help us get it done
  - ***TLA, ETLA, and Jargon*** that needs to be defined
  - Overview of **`what to type`**
  - Hands-on lab with step by step **`commands`** and `expected output.`
  - ☺*Tricky tips for experts*
  - ☠Common pitfalls

How To Get Authorized to Run Jobs

- Get the client authorization and job submission software.
  - The Open Science Grid (*OSG*) Client software from the Virtual Data Toolkit (*VDT*) is already installed on fnpcsrv1
- Get a certificate to authenticate yourself
  - All Fermilab staff already have an X.*509 certificate* from the Kerberos Certificate Authority (*KCA*)
- Become a part of an authorized grid organization.
  - All Fermilab staff and users are part of the Fermilab Virtual Organization (VO) automatically
- Find a computing facility where that organization can run
  - *FermiGrid* accepts jobs from all VO's in OSG.

Term Definitions

- *OSG:* Open Science Grid: http://www.opensciencegrid.org/
  - 72 sites mostly in the United States who share compute and storage resources with each other. Three of those sites are here at Fermilab.
- *VDT*: Virtual Data Toolkit http://vdt.cs.wisc.edu/
  - OSG-funded collection of all software needed to run on the grid. Three subsets:
  - Compute Element (CE): Software needed to set up a server to accept grid jobs.
  - Worker Node Client: Software that available on all worker node of any OSG site.
  - Client: Software that is needed to submit jobs.
- *Globus Toolkit* http://www.globus.org/
  - A wide set of services for grid job submissions, file transfer and more.
  - Included as part of the VDT
- *VO*: Virtual Organization
  - A dynamic collection of Users, Resources, and Services for sharing of Resources.
  - Examples at Fermilab: CDF, Dzero, CMS, ILC, SDSS, DES, Fermilab

Term Definitions 2

- *CA:* Certificate Authority
  - An entity that issues certificates.
- *KCA*: Kerberos Certificate Authority
  - A server that takes your Kerberos credential and issues a X.509 Certificate. Fermilab has the only one in production.
- *X.509 Certificate:*
  - Digitally signed statement from one entity (CA) saying that the public key of another entity (user) is valid. Follows the X.509 standard.
  - X.509 certificates are used to encrypt all authentication sessions for job submission and file transfer.
- *Proxy*
  - A short-lived self-contained representation of your certificate which can be used to submit jobs to the grid

How to Make Credentials for Job Submission

- On your desktop machine, get a kerberos credential
  - **`kinit -r 168h <username>@FNAL.GOV`**
- Log into a machine that has the client software on it:
  - **`ssh -l <username> fnpcsrv1.fnal.gov`**
- Source the setup file
  - **`source /usr/local/vdt/setup.sh`**
- Obtain a Fermilab KCA certificate
  - **`kx509`**
  - **`kxlist -p`**
- Get the certificate signed by the Fermilab *VOMS* server
  - **`voms-proxy-init  -noregen -voms fermilab:/fermilab`**
- Verify that the voms-proxy-init worked
  - **`voms-proxy-info -all`**
- In the next slides and Lab 1 we will go through all of these steps one by one and explain the options and the intermediate results.

Preparing to submit—sample output

```
bash-3.00$ source /usr/local/vdt/setup.sh
bash-3.00$ kx509
bash-3.00$ kxlist -p
Service kx509/certificate
 issuer= /DC=gov/DC=fnal/O=Fermilab/OU=Certificate Authorities/CN=Kerberized CA
 subject= /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C. Timm/UID=timm
 serial=7E6C63
 hash=03c202fc
bash-3.00$ voms-proxy-init -noregen -voms fermilab:/fermilab
Cannot find file or dir: /home/condor/execute/dir_11128/userdir/glite/etc/vomses
Your identity: /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C. Timm/USERID=timm
Cannot find file or dir: /home/condor/execute/dir_11128/userdir/glite/etc/vomses
Contacting  voms.fnal.gov:15001
[/DC=org/DC=doegrids/OU=Services/CN=http/voms.fnal.gov] "fermilab" Done
Creating proxy ................................ Done
Your proxy is valid until Tue Feb 26 07:41:27 2008
```

- Comments—The warning about missing /home/condor directory is routine
- -voms fermilab:/fermilab is a Fully Qualified Attribute Name (*FQAN*), see handout for details

How did you know it worked?

```
bash-3.00$ voms-proxy-info -all
WARNING: Unable to verify signature! Server certificate possibly not installed.
Error: Cannot find certificate of AC issuer for vo fermilab
subject   : /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C.
Timm/USERID=timm/CN=proxy
issuer    : /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C. Timm/USERID=timm
identity  : /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C. Timm/USERID=timm
type      : proxy
strength  : 512 bits
path      : /tmp/x509up_u2904
timeleft  : 10:41:35
=== VO fermilab extension information ===
VO        : fermilab
subject   : /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C. Timm/USERID=timm
issuer    : /DC=org/DC=doegrids/OU=Services/CN=http/voms.fnal.gov
attribute : /fermilab/Role=NULL/Capability=NULL
timeleft  : 10:41:35
```

Error message about “cannot find certificate” can be ignored

Why doesn't voms-proxy-init work?

```
bash-3.00$ voms-proxy-init -noregen -voms cms:/cms
Cannot find file or dir: /home/condor/execute/dir_11128/userdir/glite/etc/vomses
Your identity: /DC=gov/DC=fnal/O=Fermilab/OU=People/CN=Steven C. Timm/USERID=timm
Cannot find file or dir: /home/condor/execute/dir_11128/userdir/glite/etc/vomses
Contacting  lcg-voms.cern.ch:15002 [/DC=ch/DC=cern/OU=computers/CN=lcg-
voms.cern.ch] "cms" Failed
Error: cms: User unknown to this VO.
Trying next server for cms.
Contacting  voms.cern.ch:15002 [/DC=ch/DC=cern/OU=computers/CN=voms.cern.ch]
"cms" Failed
Error: cms: User unknown to this VO.
None of the contacted servers for cms were capable
of returning a valid AC for the user.
```

- You might not be a member of the VO (see error message above)
- You might be requesting a role that you aren't authorized to be.
- Check by going to *VOMS* server https://voms.fnal.gov:8443/voms/fermilab
- **`voms-proxy-init -debug`** is your friend (Could be missing a vomses file.)
- To join a VO that you're not in now, use *VOMRS* to request membership.
- *VOMS* server might be down.

Term Definitions 3

- *VOMS*—Virtual Organization Management Service
  - All Virtual Organizations use this to certify that a member is part of their VO and has certain rights and privileges
- *VOMRS*—Virtual Organization Membership Registration Service
  - A frontend to VOMS that handles policy signing, expirations, adding extra certificates, group and role management, and more.
- *FQAN*—Fully Qualified Attribute Name
  - The combination of group and role for the user

Authorization and Credentials: Lab 1

- See the handout for Lab 1.
- Use the kx509/kxlist -p /voms-proxy-init sequence to get a good voms proxy.
- Show the instructor when you think you have correct output of voms-proxy-info -all.
- Try the other examples after section A and B if you have more time and are waiting for others.

Grid job submission in English

- There is a submission machine and a compute element (*CE*).
  - In this example, fnpcsrv1=submission machine, fngp-osg=compute element
- Client side authenticates to the compute resource
  - Using your certificate and the machine's certificate to make a SSL connection
- The executable and input files are transferred to the compute resource
  - By opening an https: connection
- The executable is submitted to the batch system on the compute resource
  - Using the *GRAM* interface
- When the job completes, the output files are transferred back.
  - Again using an https: port

- *GRAM*=Globus Resource Access Manager

Test submit: Globus-job-run

- Example
  - `Globus-job-run fngp-osg.fnal.gov:2119/jobmanager-fork /usr/bin/id`
- Structure of the example:
  - Host:port to submit the job to.
    - 2119 is the default port and can be omitted
  - Which jobmanager to use?
    - Jobmanager-fork is the default. This runs jobs directly on the CE. Production jobs should be jobmanager-condor or jobmanager-pbs.
  - Command to use
    - This example will run the /usr/bin/id that's already on the remote machine.
- Comments
  - Globus-job-run should be used only for diagnostic purposes
  - One daemon per globus-job-run is launched on the remote machine and stays running until it exits—or sometimes hangs.

Test transfer: globus-url-copy

- Globus-url-copy is the command-line client for GRIDFTP
- Example:
  - ```
    globus-url-copy file://${HOME}/fermigrid201/lab2/foo \
    gsiftp://fngp-osg.fnal.gov/grid/data/foo.${USER}
    ```
- Comments:
  - Globus-url-copy is for small files and light testing
  - Can be used for big files too, but with a management software like srmcp.
  - In the above example, the environment variables are evaluated on submit machine
  - Works to go to compute elements or storage elements
    ```

Grid Submission: globus-job-run and globus-url-copy

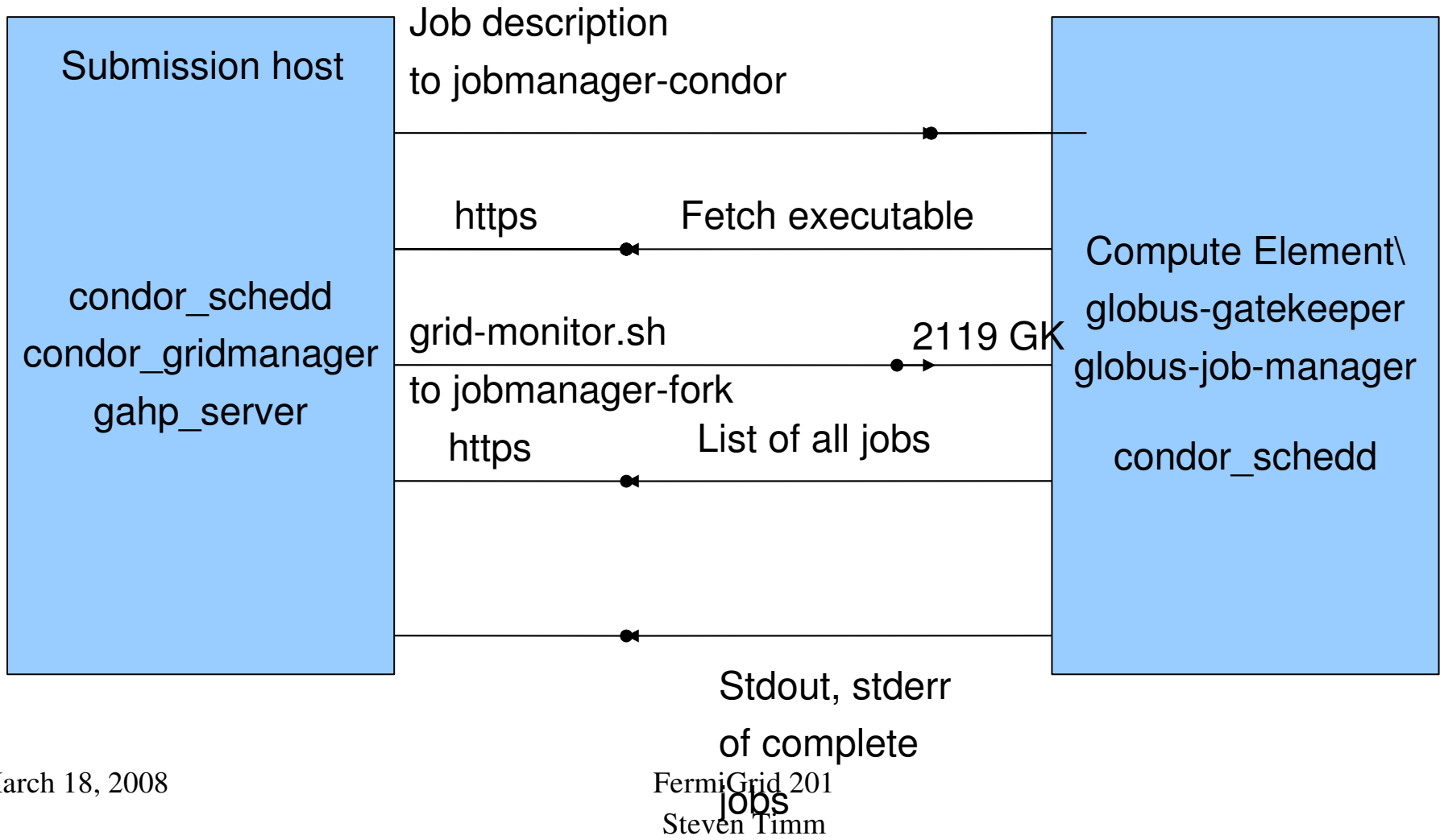
Lab 2

- Execute the following sequence:
- globus-job-run fngp-osg.fnal.gov:2119/jobmanager-fork /usr/bin/id
- globus-url-copy file:///${HOME}/helloworld.sh gsiftp://fngp-osg.fnal.gov/grid/data/helloworld.sh.${USER}
- globus-job-run fngp-osg.fnal.gov:2119/jobmanager-fork /bin/chmod 755 \ /grid/data/helloworld.sh.${USER}
- globus-job-run fngp-osg.fnal.gov:2119/jobmanager-fork \ /grid/data/helloworld.sh.${USER}

Condor submission concepts in English

- Condor is comprehensive batch system and grid submission software
- Grid submission client components are called Condor-G
- Have to install all of Condor to use the Condor-G clients.
- Condor-G runs on the submission host and
  - Transfers your executable and input files to remote compute element and gets it started
  - Monitors the status of the job every minute to see if it is done
  - Transfers the files back when the job is over.
- On the client machine, the condor_schedd keeps track of all jobs and spawns a condor_gridmanager to send the jobs to the grid

Condor-G Job Submission



Condor submission—simple example

```
universe = grid
GridResource = gt2 fngp-
osg.fnal.gov/jobmanager-condor
executable = recon1
transfer_output = true
transfer_error = true
transfer_executable = true
stream_output = false
stream_error = false
log = grid_recon1.log.$(Cluster).$(Process)
notification = NEVER
output=grid_recon1.out.$(Cluster).$(Process)
error = grid_recon1.err.$(Cluster).$(Process)
globusrsl = (jobtype=single)(maxwalltime=999)
queue
```

Grid universe for all jobs

type gt2 refers to version 2 of Globus

recon1 is a binary that will run for 3 minutes

Annotated version in the examples

To submit it:

condor_submit grid_recon1

Transferring input and output files

```
bash-3.00$ more fngp-osg-gridsleep-fourargs
Universe = grid
remote_initialdir = /grid/data/foo
GridResource = gt2 fngp-osg/jobmanager-condor
executable = gridsleep.sh
# Old style of condor arguments
arguments = one two three four
transfer_output = true
transfer_error = true
transfer_executable = true
stream_output = False
stream_error = False
should_transfer_files = YES
when_to_transfer_output = ON_EXIT_OR_EVICT
transfer_input_files = foo
transfer_output_files = bar
log = gridsleep.log.$(Cluster).$(Process)
notification = NEVER
output = gridsleep.out.$(Cluster).$(Process)
error = gridsleep.err.$(Cluster).$(Process)
globusrsl = (condorsubmit=(requirements
'Disk>5000'))
queue 1
```

condor_q and condor-q -globus

```
[root@fnpcsrv1 ~]# condor_q timm
 ID      OWNER            SUBMITTED     RUN_TIME ST PRI SIZE CMD
1704117.0   timm            3/16 21:13   0+00:00:00 I  0   0.0  gridsleep_files.sh
[root@fnpcsrv1 ~]# condor_q -globus timm
 ID      OWNER          STATUS  MANAGER  HOST                EXECUTABLE
1704117.0   timm          UNSUBMITTED condor   fermigrid1.fnal.go  /home/timm/gridsle
[root@fnpcsrv1 ~]# condor_q -globus timm
 ID      OWNER          STATUS  MANAGER  HOST                EXECUTABLE
1704117.0   timm          PENDING condor   fermigrid1.fnal.go  /home/timm/gridsle
[root@fnpcsrv1 ~]# condor_q -globus timm
 ID      OWNER          STATUS  MANAGER  HOST                EXECUTABLE
1704117.0   timm          ACTIVE condor   fermigrid.fnal.go /home/timm/gridsleep
```

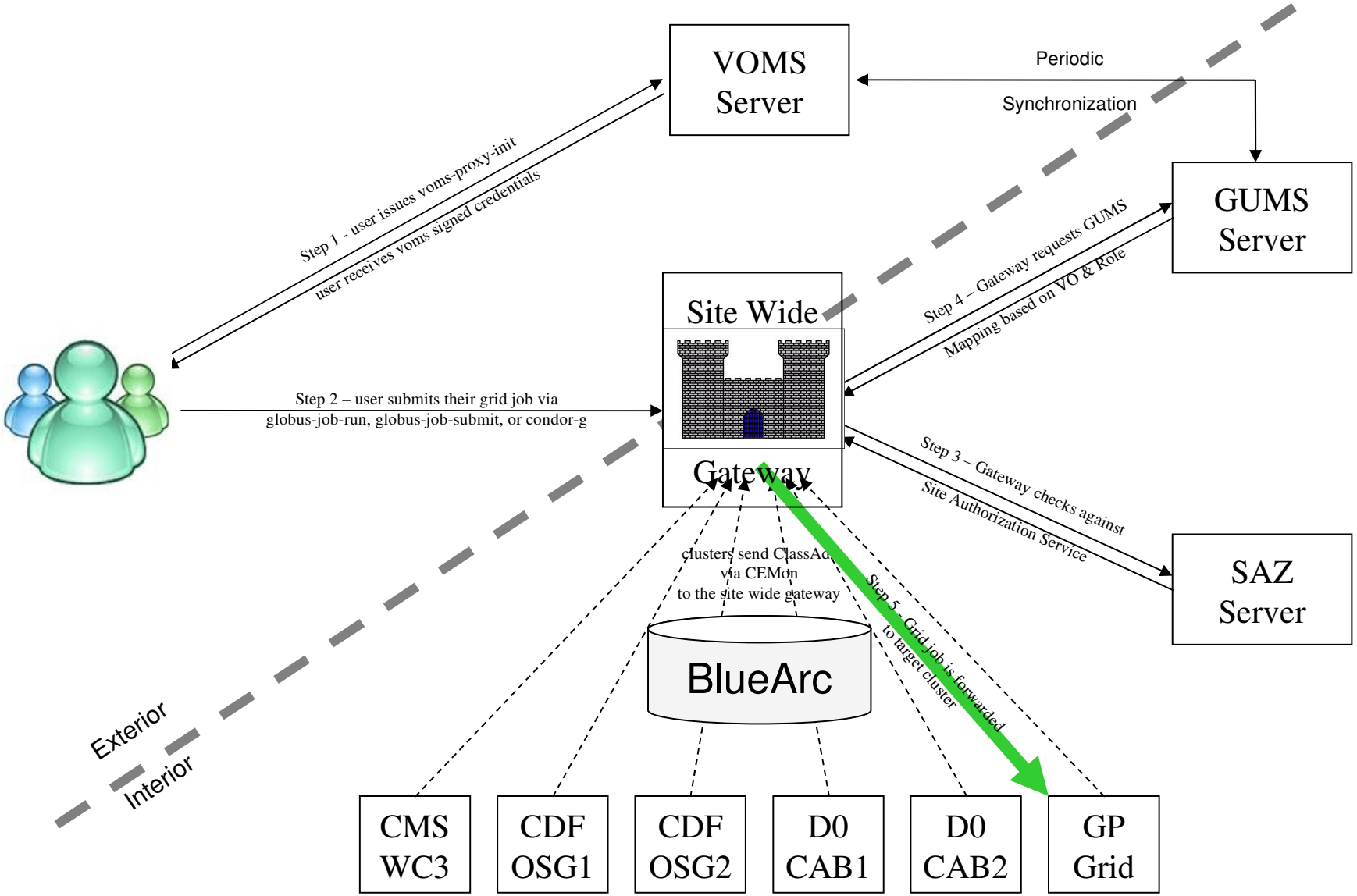
Condor_q -globus shows the status of the job as it moves through the grid.

Shows “ACTIVE” once the job starts running remotely.

Grid Submission: condor_submit--Lab 3

- Submit the jobs grid_recon1 and fngp-osg-gridsleep-fourargs
- Monitor their progress with condor_q and condor_q -globus
- Record any errors

FermiGrid - Current Architecture



Data movement on FermiGrid

- Most grid sites have a disk area for applications, where certain users are allowed to install application software.
  - On FermiGrid this area ($OSG_APP) is /grid/app, NFS mounted from Bluearc to all GP, CDF, D0 clusters
- Most grid sites have a disk area for data, where users are allowed to put data so that it is accessible to worker nodes
  - On FermiGrid this area ($OSG_DATA) is /grid/data, NFS mounted from Bluearc to all GP, CDF, D0 clusters.
- Many grid sites also have an SRM-based Storage Element
  - (see this afternoon's class FermiGrid 202 for details)
  - Fermilab's is srm://fndca1.fnal.gov:8443/
- There is also a scratch area per job ($OSG_WN_TMP) on local worker node disk.
  - On FermiGrid this directory is usually /local/stage1—but see the examples for detecting it correctly every time.

Discovering compute resources and directories, OSG

- Look them up in advance via *VORS* http://vors.grid.iu.edu/
  - *VORS*=Virtual Organization Resource selector
- Or detect them in your job when you get there
- FermiGrid also has regular testing for sites that accept the Fermilab VO at

  http://fermigrid.fnal.gov/monitor/fermigrid0-fermilab-vo-production-monitor-summary.html
- Two alternate data flow models
  - Push files to /grid/data in advance of the job
  - Get to the worker node and pull files there, once you get there.

Data Flow Models

Push Model

Find $DATA from VORS

Copy data to remote site $DATA

Analyze data from $DATA

Copy results back from remote $DATA

Pull model

Submit job, it starts running on WN

Source OSG_GRID/setup.sh to find $DATA and $WN_TMP

Pull data straight to worker node disk

Push result from worker straight to remote site

FermiGrid and OSG, Lab 4

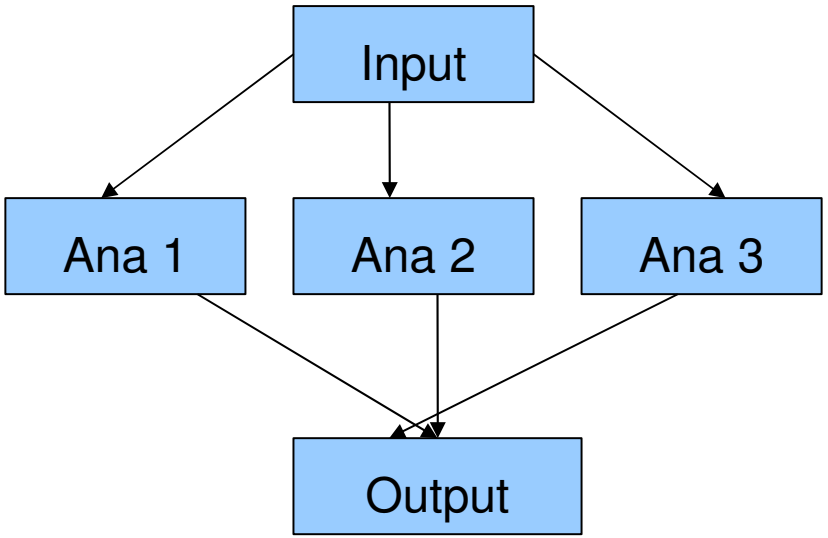
- Submit sample single job from tarball that auto-detects OSG_DATA and OSG_APP and OSG_WN_TMP, and uses them.
- Go to VORS and FermiGrid and find the information for one other OSG site.
- Modify the submit file to send the test job to that OSG site instead.

Globus RSL

- *RSL*=Resource Specification Language
- The way to communicate requirements to the remote batch system
- Can be used to set memory, wall time, processor type, architecture, and more. We have examples http://fermigrid.fnal.gov/gpgrid/examples
- jobtype=single—needed for most PBS sites, can use anywhere
- queue=xxxxx—needed for most PBS sites

Condor DAGman

- DAG=Directed Acyclic Graph
- Used to show dependencies—to make one job not start until its predecessor is completed.
- Example is provided in the example tarball, we will go through it if we have time
- Pegasus as automated DAG maker
- **`condor_submit_dag ex2.dag`**



Using DOEGrids Certificates

- Why get a DOEGrids cert? (see http;//security.fnal.gov/pki for full explanation)
  - More widely accepted
  - Can load into browser long term
  - Can use to sign E-mail
- Store your DOEGrids cert and private key—on some non-network-mounted disk.
- If a certificate is compromised, you revoke it by contacting DOEGrids.
- Open Science Enclave policy requires you protect cert and proxy.

Monitoring of Grid Jobs

- Globus GRAM is meant to hide the remote batch system details from the submitting host. It is **very** good at this.
- condor_q
- condor_q -globus
- condor_q -held
- Userlog
- CondorView
- Remote condor_q
  - condor_q -name fngp-osg -pool fnpcsrv2
  - condor_q -name fcdfosg2 -pool fcdfcm2
  - condor_q -name cmsosgce3 -pool cmssrv14.fnal.gov

condor_q -held

```
[root@fnpcsrv1 ~]# condor_q -held
 ID       OWNER            HELD_SINCE HOLD_REASON
1674997.0   greenc           3/11 09:35 Globus error 12: the
connection to the serv
1702530.9   carneiro         3/15 02:12 Globus error 131: the
user proxy expired(j
1703617.0   rubin            3/15 16:29 Globus error 5: the
executable does not exi
1703626.0   rubin            3/15 17:21 Globus error 10: data
transfer to the serve
1703631.0   rubin            3/15 17:22 Globus error 3: an I/O
operation failed
```

Condor_q -held tells you when the job was held, and why.

First thing to try is to **condor_release** the job

To remove, first **condor_rm**,

If that doesn't work, **condor_rm -forcex.**

Condor Userlog

To contact remote admin on a failed job you need three things:

1) Timestamp when it happened
2) Globus job id
3) Name of machine you are submitting from.

```
[root@fnpcsrv1 ~]# more
/minos/data/minfarm/condor_log/mca_n13037637_0027_L010185N_D04.0.269
79.1.log
000 (1704008.000.000) 03/16 08:29:19 Job submitted from host:
<131.225.167.44:61501>
017 (1704008.000.000) 03/16 08:30:29 Job submitted to Globus
    RM-Contact: fermigrid1.fnal.gov/jobmanager-condor
    JM-Contact: https://fermigrid1.fnal.gov:49043/28668/1205674220/
    Can-Restart-JM: 1
027 (1704008.000.000) 03/16 08:30:29 Job submitted to grid resource
    GridResource: gt2 fermigrid1.fnal.gov/jobmanager-condor
    GridJobId: gt2 fermigrid1.fnal.gov/jobmanager-condor
https://fermigrid1.fnal.gov:49043/28668/1205674220/
001 (1704008.000.000) 03/16 08:37:55 Job executing on host: gt2
fermigrid1.fnal.gov/jobmanager-condor
012 (1704008.000.000) 03/16 09:19:00 Job was held.
        Globus error 10: data transfer to the server failed
        Code 2 Subcode 10
```

FNAL_GPFARM Condor Pool User Statistics for Month - Mozilla Firefox

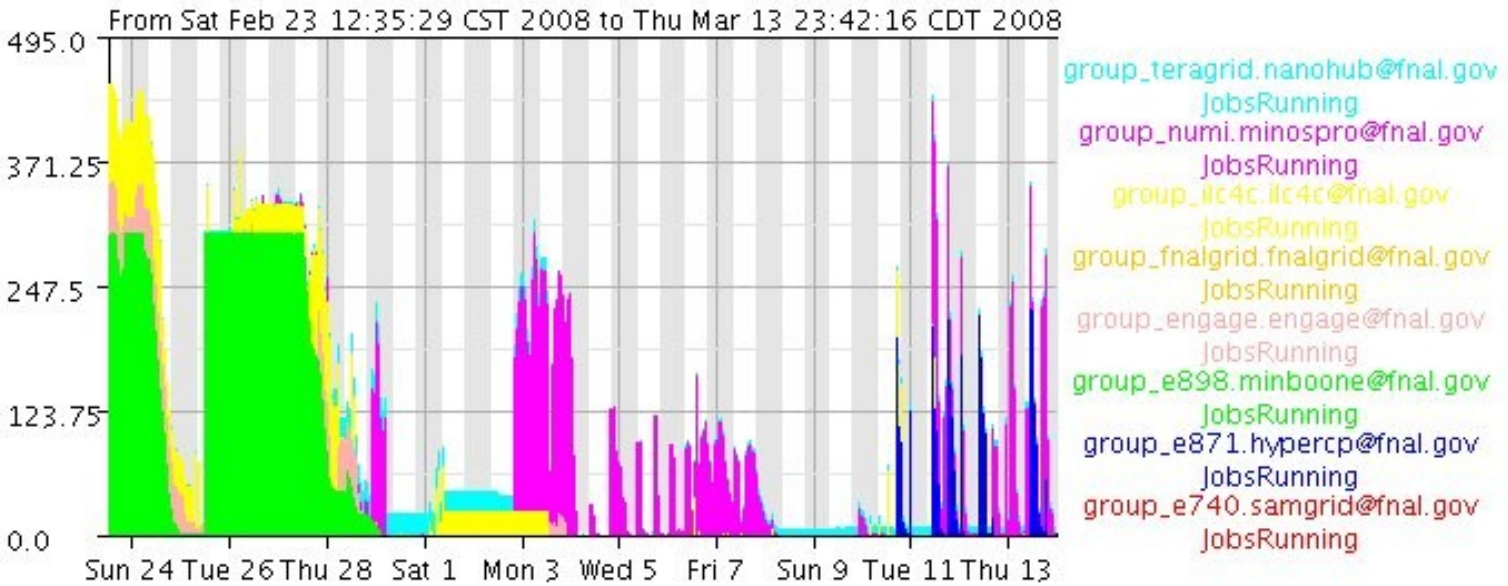
File Edit View Go Bookmarks Tools Help

http://fnpcsrv1.fnal.gov/condorview/viewdir/UserMonth.html

Go

Fermi Linux Fermilab FermiLinux Linux Distros

FNAL_GPFARM Condor Pool User Statistics for Month



Configure... Zoom In Zoom Out Reset About

[Graph Hints: The Y-axis is number of jobs, the X-axis is time. When graph finishes updating, press "Configure.." to view different Architecture or State data. Also, you can use the mouse to draw a rectangle on the graph and then press "Zoom In". Press "Reset" to center/resize the data after Configure or when done zooming. Nighttime shows up on graph background as grey.]

User	Total Allocation Time (Hours)	JobsRunning Average	JobsIdle Average	JobsRunning Peak	JobsIdle Peak
Total	121916.2	261.4 (88.9%)	97.5 (11.1%)	810.0 (100.0%)	1194.0 (70.7%)
[illegible]	[illegible]	2.9	0.0	5.0	0.0

Applet TimeGraph started

Problem diagnosis—Globus Errors

- Globus error 3—I/O error when transferring file
- Globus error 7—authentication, at Fermilab usually a problem with SAZ or GUMS
- Globus Error 10—failure to transfer file, means something is out of quota somewhere.
- Globus Error 12—can't contact the gatekeeper, either it is down or you typed the hostname wrong.
- Globus error 17—either the executable isn't there or there is something wrong with the batch system.
- Globus error 31—failed to cancel the job
- Globus error 93—gatekeeper failed to find the requested service. Probably you requested a jobmanager that wasn't there
- Globus error 155—failure to stage out—happens when proxy expires before end of job
- http://www.cs.wisc.edu/condor/manual/v7.0/Appendix_B_Magic.html

Problem Diagnosis, other errors

```
bash-3.00$ globus-job-run d0cabosg1/jobmanager-condor /usr/bin/id
ERROR: proxy does not exist
Syntax : globus-job-run {[-:] <contact string> [-np N] <executable> [<arg>...]}...
Use -help to display full usage
```

Proxy isn't there. You have to voms-proxy-init

```
Error: Could not establish authenticated connection with the server.
GSS Major Status: Authentication Failed
GSS Minor Status Error Chain:
init.c:globus_gss_assist_init_sec_context:277:
Error during context initialization
init_sec_context.c:gss_init_sec_context:190:
SSLv3 handshake problems
globus_i_gsi_gss_utils.c:globus_i_gsi_gss_handshake:889:
Unable to verify remote side's credentials
globus_i_gsi_gss_utils.c:globus_i_gsi_gss_handshake:862:
SSLv3 handshake problems: Couldn't do ssl handshake
OpenSSL Error: s3_clnt.c:842: in library: SSL routines, function SSL3_GET_SERVER_CERTIFICATE:
certificate verify failed
globus_gsi_callback.c:globus_gsi_callback_handshake_callback:531:
Could not verify credential
globus_gsi_callback.c:globus_i_gsi_callback_cred_verify:729:
Could not verify credential
globus_gsi_callback.c:globus_i_gsi_callback_check_revoked:1031:
Invalid CRL: The available CRL has expired
```

This means that your Certificate Revocation List is old, need to fix.

Problem diagnosis: UPS/UPD, expired proxies

- Avoid setting up UPS/UPD products in a grid job if possible
- UPS/UPD products bring along obsolete versions of perl and python which are not compatible with some grid utilities.
- If proxy expires in mid-job, can still rescue job.
  - condor_q -held shows the error that the proxy is expired
  - Renew your proxy by normal methods
  - Then condor_release the job.

Intro to Globus Web Services

- Globus toolkit is moving from clunky and slow and old perl scripts to clunky and slow and new Java applications.
- Instead of globus-job-run, globusws-run.
- Deployed everywhere on FermiGrid and most of the OSG but the monitoring and information systems haven't caught up yet.
- Old way will stay around for quite a while.
- More web services jobs come later.

Advanced Topics--Lab 5

- Submit sample DAG
- Monitor by using condor_q to fngp-osg
- If you have a doegrids cert, voms-proxy-init and submit a job with that.
- Submit a Globus-WS job with globusrun-ws
- `Globusrun-ws -submit -F fngp-osg.fnal.gov:9443 -s -J -c /usr/bin/id`

End of class

- Be sure to issue the “kdestroy” command to destroy your kerberos credentials on the test machine
- Make sure you filled in an evaluation
- Thanks for coming!
- Further questions can always go to http://helpdesk.fnal.gov
- With problem “Grid”, type “Fermilab Sup. Center”, item “FermiGrid”