Getting Started with Grid Computing in STAR

Eric Hjort, LBNL

STAR Collaboration Meeting

August 13, 2003

Purpose of the Talk

- Encourage STAR users to get started on Grid Computing:
  - **It's not hard to do some simple, useful things**
  - **Soon the STAR job scheduler will be grid-enabled**
  - **Important to start building grid expertise among STAR users**
- Provide a step-by-step tutorial:
  - **Obtain a DOE Grid certificate**
  - **Get your certificate installed and working**
  - **Examples of a few simple grid commands**

A Couple Definitions…

- What is Globus? The Globus Project produces the Globus Toolkit, which has become the standard for software often called grid middleware.

- What is PPDG? PPDG is the Particle Physics Data Grid, a collaboration between 4 computer science projects and 6 physics experiments, including STAR.

Getting a DOE Grid Certificate

- Start at http://www.doegrids.org
- Follow the links “How to Request Certificates”, then “How to Request Personal Certificates”.
- Follow instructions, and provide the following information:

  Affiliation = PPDG

  Sponsor = your supervisor

  Additional Comments = STAR experiment
- Click “Submit” to get your certificate issued.

Importing your Certificate into your Browser

- Once your certificate has been issued you will receive an email containing instructions and a link.
- Follow the link and click on the “Import Your Certificate” button. This will import your certificate into your browser.
- You can see your certificate:

  IE: Tools->Internet options->Content

  Netscape: Edit->Preferences->Privacy&Security->Certificates

Exporting your certificate

- Continuing with the instructions on the doegrids.org page, the next step is to export your certificate from your browser to your computer (look for an “export” button where you saw your certificate).
- This will result in the creation of a file on your computer, possibly with an extension of .p12, but depending on your browser possibly something else like .pkcs12.

Generating your key pair

- It is necessary to install your certificate at any facility where you want to do grid commands, e.g., grid-proxy-init. To do this:
- Copy your .p12 file from your computer to your account at PDSF or RCF.
- Create a .globus subdirectory of your home directory
- Generate your key pair:

  openssl pkcs12 –in YourCert.p12 –clcerts –nokeys –out $HOME/.globus/usercert.pem

  openssl pkcs12 –in YourCert.p12 –nocerts -out $HOME/.globus/userkey.pem
- Reset the permissions on the userkey.pem file you just created: "chmod 600 userkey.pem".
- Your certificate is now installed!

Updating grid-mapfiles

- The final step before you can use grid tools is to get the grid-mapfile updated at each of STAR's grid-enabled sites (RCF and PDSF at present).
- The grid-mapfile maps your certificate to your local account at that site.
- The grid-mapfile is updated by local site administrators and the procedure is site-dependent.
- At RCF you can make a request through the BNL bugzilla system http://bugzilla.bnl.gov:
  - **Click on Enter a new bug report**
  - **Click on Grid Related and you will go to the BNL Grid Bug report system.**
  - **Select "user registration" from the product list**
  - **Enter your account information and the subject and issuer lines from your usercert.pem file**

Updating grid-mapfiles (cont'd)

- At PDSF grid-mapfile updates are handled through NIM (NERSC Information Management)
  - **Go to NIM: https://nim.nersc.gov/**
  - **Login.**
  - **Under "My Stuff" select "My Contact Info"**
  - **Copy the subject and issuer info from your usercert.pem file into the corresponding fields (down at the bottom)**
- The PDSF grid-mapfiles should be updated with your information within 24 hours

Are we there yet???

- If you can do a grid-proxy-init, then YES!
- Test it at a facility where you have the .pem files in your .globus directory.
- At RCF:
  - **Login to stargrid02.rcf.bnl.gov**
  - **"grid-proxy-init"**
- At PDSF:
  - **Login to pdsfint.nersc.gov as usual**
  - **"module load globus"**
  - **"grid-proxy-init"**
- If grid-proxy-init accepts your GRID pass phrase your proxy will be created:

Your identity: /O=doesciencegrid.org/OU=People/CN=Eric Hjort 8713

Enter GRID pass phrase for this identity:

Creating proxy .... Done

Your proxy is valid until Tue Aug 12 05:16:53 2003

Quick summary in 6 steps

- Request a certificate from doegrids.org
- When notified, import certificate to your browser
- Export your certificate from your browser to your computer
- Copy your certificate from your computer to each grid-enabled facility that you plan to work at
- generate a key pair at each facility
- Get the grid-mapfiles updated at all grid-enabled STAR computing facilities
- Test with grid-proxy-init

Note: This process can be confusing, especially the first time you try it. There are numerous options (usually the defaults are OK) and various passwords involved so it's a good idea to take notes as you try it.

Simple grid commands

- grid-proxy-init:
  - **Initializes your grid proxy. Default lifetime 12 hours.**
  - **Use option "-hours N" to increase to N hours.**
- grid-proxy-info:
  - **Shows info on your proxy, including time until expiration.**
- PDSF user copies a file RCF->PDSF:
  - **globus-url-copy -vb gsiftp://stargrid02.rcf.bnl.gov/star/u/hjort/test.txt file:/auto/u/hjort/test.txt**
  - **-vb is verbose mode**
  - **gsiftp: refers to a file in a remote location and stands for Globus Security Infrastructure File Transfer Protocol. In this example it is used in the source URL.**
  - **file: refers to a local file, in this case the destination URL**

globus-job-run

- globus-job-run: do a command on a remote site.
  - **Example: globus-job-run stargrid02.rcf.bnl.gov /bin/ls /star/u/griduser**
  - **This command, issued from PDSF, would list the contents of the directory /star/u/griduser at RCF**
  - **Note that the path is needed for the executable (ls).**
- -s option: execute my script (located at PDSF) at RCF:
  - **Example: globus-job-run stargrid02.rcf.bnl.gov –s gridjob.pl**
  - **Convenient because I can edit my script locally but execute it remotely**
  - **Can also use –stdin and –stdout options to specify stdin and stdout**

globus-job-submit and globusrun

- globus-job-submit
  - **Similar to globus-job-run but with a batch interface**
  - **Doesn’t yet support the staging (-s) option.**
- globusrun
  - **Allows user to use Globus Resource Specification Language (RSL) to specify more complex jobs.**
  - **globus-job-run and globus-job-submit are wrappers around globusrun**

Resources

- The Globus website: http://www.globus.org
- The Globus 2.4 user's Guide: http://www-fp.globus.org/gt2.4/admin/guide-user.html (STAR does not use Globus 3 yet)
- Iwona Sakrejda's "Introducing Grid Services at NERSC" is at http://hpcf.nersc.gov/training/classes/2003/Jan.23/#sak

  Iwona provides a much more detailed description of things discussed briefly in this tutorial.

Summary

- This tutorial is meant to supply just enough information to get you started on grid computing in STAR
- Getting your certificate can be confusing, and the details of how to do it can depend on things such as what browser you use, etc.
- Once you are set up, however, there is little or no maintenance – just remember your pass phrase
- Anybody that does STAR computing, especially people who use both RCF and PDSF, is encouraged to spend an hour or two trying to get started on the grid.