WN Hands-on Session

www.see-grid-sci.eu

Regional SEE-GRID-SCI Training for Site Administrators
Institute of Physics Belgrade
March 5-6, 2009

Miloš Ivanović
AEGIS04-KG
Faculty of Science, University of Kragujevac
Serbia
mivanovic@kg.ac.rs

The SEE-GRID-SCI initiative is co-funded by the European Commission under the FP7 Research Infrastructures contract no. 211338

Introduction

- OS installation & tunung
- Repository adjustment
- Java installation
- File system import/export
- gLite middleware packages installation
- SSH configuration
- gLite configuration
- Post-istallation tips
- Installed system testing
- WN replication

OS installation & configuration

- **Newest Scientific Linux series 4** (currently 4.7) should be installed (not SL5)
- Both **32-bit** and **64-bit** distribution are supported by glite-WN and MPI_WN so far
- We have chosen to **install all base packages** from all 5 SL4.7 CDs, then remove unnecessary
- Packages with great chances not to be used should be removed to speed up future software updates, i.e. openoffice.org
- **Remove all LAM and OPENMPI packages**, we'll be using MPICH
- **Remove java-1.4.2-sun-compat** package!
- Virtual environment is a possible solution

Further OS tuning

- **Adjust services/daemons** started at the boot time
  - it is recommended to change the **default runlevel to 3** in /etc/inittab
  - **disable yum auto-update**, since this may bring trouble when new gLite updates appear
  - If you install MPI_WN, it is suggested to disable SELINUX by replacing “SELINUX=enforcing” with line “**SELINUX=disabled**” in the file /etc/selinux/config
- **Configure NTP** service
  - Example of configuration file /etc/ntp.conf can be found on http://glite.phy.bg.ac.yu/GLITE-3/ntp.conf
  - touch /etc/ntp.drift /etc/ntp.drift.TEMP
  - chown ntp.ntp /etc/ntp.drift /etc/ntp.drift.TEMP
  - chkconfig ntpd on

Repository adjustment

- **DAG repo** should be enabled by changing "enabled=0" into "enabled=1" in /etc/yum.repos.d/dag.repo. After package installation takes place, DAG repo should be disabled
- **Base SL repos must be PROTECTED**, disallowing DAG packages to replace them! Add line “protect=1” to /etc/yum.repos.d/sl.repo and /etc/yum.repos.d/sl-errata.repo

SCL repository

- In order to install glite-WN with MPI support, following repo configuration files need to be downloaded into /etc/yum.repos.d:
  - http://rpm.scl.rs/yum.conf/scl-glite-TORQUE_client.repo
  - http://rpm.scl.rs/yum.conf/scl.repo
  - http://rpm.scl.rs/yum.conf/scl-glite-WN.repo
  - http://rpm.scl.rs/yum.conf/scl-lcg-ca.repo
  - http://rpm.scl.rs/yum.conf/scl-glite-MPI_utils.repo
  - http://rpm.scl.rs/yum.conf/scl-jpackage.repo
- Java packages are already present in the repo, but newest version can be installed following instructions on the next slide.

Java installation

- Use latest Java 1.5! Follow advice from:
  https://twiki.cern.ch/twiki/bin/view/EGEE/GLite31JPackage
  or
  http://wiki.egee-see.org/index.php/SL4_WN_glite-3.1

- To install it, it is necessary to go to SUN's Java web page and download JDK 5.0 Update 15. We used "Linux self-extracting file" jdk-1_5_0_15-linux-i586.bin in order to make java-1.5.0-sun-1.5.0.15-1jpp.i586.rpm and java-1.5.0-sun-devel-1.5.0.15-1jpp.i586.rpm packages, as suggested in Steve Traylen's guide. To make and install those two packages, do the following:

- ```
  rpm --import http://www.jpackage.org/jpackage.asc
  mkdir -p ~/redhat/BUILD ~/redhat/SOURCES ~/redhat/SPECS ~/redhat/RPMS/i586 ~/
  redhat/SRPMS
  cat <<EOF > ~/.rpmmacros
  %_topdir    $HOME/redhat
  %packager        Firstname Lastname <firstname.lastname@example.org>EOF
  rpm -Uvh http://mirrors.dotsrc.org/jpackage/1.7/generic/non-free/SRPMS/java-
  1.5.0-sun-1.5.0.15-1jpp.nosrc.rpm
  cp jdk-1_5_0_15-linux-i586.bin ~/redhat/SOURCES
  ```

- ```
  rpmbuild -ba ~/redhat/SPECS/java-1.5.0-sun.spec
  rpm -Uvh ~/redhat/RPMS/i586/java-1.5.0-sun-1.5.0.15-1jpp.i586.rpm
  ~/redhat/RPMS/i586/java-1.5.0-sun-devel-1.5.0.15-1jpp.i586.rpm
  ```

File system import/export

- Application software filesystem
  - All WNs must have **shared application software filesystem** where VO SGMs (software grid managers) will install VO-specific software.
  - If it's supposed to be located on SE, following (or similar) line must be appended to /etc/exports
    ```
    /opt/exp_soft 147.91.12.0/255.255.255.0(rw,sync,no_root_squash)
    ```
  - If you want to map application software filesystem from other node (usually SE), append this line to /etc/fstab:
    ```
    se.csk.kg.ac.yu:/opt/exp_soft /opt/exp_soft  nfsdefaults    0 0
    ```
    Do not forget to create /opt/exp_soft dir on each WN!
- **Shared /home** filesystem:
  - In order to **provide appropriate MPI support**, entire /home must be shared among WNs.
  - Procedure is equal to procedure for app. soft. filesystem

gLite software installation

- Host certificate is not necessary on WN
- gLite software binaries, libraries and other stuff are organized using meta-package paradigm. In order to install necessary packages for WN node with MPI support, following packages must be installed:
  - glite-WN
  - glite-TORQUE_client
  - glite-MPI_utils

- Due to **temporary packaging inconsistency** in glite-MPI_utils described in link, YUM command line should look like:
  - `yum install glite-WN glite-TORQUE_client glite-MPI_utils torque-2.1.9-4cri.slc4 torque-client-2.1.9-4cri.slc4 torque-mom-2.1.9-4cri.slc4.i386 --disablerepo=jpackage17-generic`
  - `yum --exclude=torque-mom --exclude=torque-client --exclude=torque update`

SSH configuration

- SSH must allow **hostbased authentication** between CE and WNs, as well as among WNs each other
- This is especially important if grid site supports **MPI**
- Helper script available in gLite can be found at `/opt/edg/sbin/edg-pbs-knownhosts`
- Script configuration can be adjusted in `/opt/edg/etc/edg-pbs-knownhosts.conf`
- Put all relevant FQDNs into **`/etc/ssh/shosts.equiv`**
- This is standard procedure for hostbased SSH
- Identical procedure applies to all WNs

gLite configuration

- All grid sevices must be configured properly using YAIM tool. Official info available at https://twiki.cern.ch/twiki/bin/view/LCG/YaimGuide400
- Templates for input YAIM files can be taken from https://viewvc.scl.rs/viewvc/yaim/trunk/?root=seegrid
- Since YAIM is mainly a set of bash scripts, bash-like syntax must be used in input files
- Required input files are:
  - site-info.def
  - users.conf
  - wn-list.conf
  - groups.conf
  - directory vo.d with one file per VO
- YAIM config. files **must not be readable for users**!

gLite configuration

- **site-info.def**
  - Main configuration input source
  - Contains proper paths to all other configuation files
- **users.conf**
  - Defines UNIX pool users for each Virtual Organization
  - Helpful script at http://glite.phy.bg.ac.yu/GLITE-3/generate-pool-accounts-AEGIS-v4
  - Example:
    ./generate-pool-accounts-AEGIS-v4 seegrid 20000 seegrid 2000 200 10 10 >> users.conf
- **groups.conf**
  - Defines groups per VO, template can be employed as is.
- **wn-list.conf**
  - Simple list of FQDNs of available Worker Nodes
- **vo.d/**
  - Directory containing a file per each supported VO.

gLite configuration - MPI

- In case of MPICH support, site-info.def file should contain lines:

```
MPI_MPICH_ENABLE="yes"
MPI_MPICH_PATH="/opt/mpich-1.2.7p1/"
MPI_MPICH_VERSION="1.2.7p1"
MPI_MPICH_MPIEXEC="/opt/mpiexec-0.82/bin/mpiexec"
MPI_SSH_HOST_BASED_AUTH="yes"
MPI_SHARED_HOME="yes"
```

- Following http://wiki.egee-see.org/index.php/SEE-GRID_MPI_Admin_Guide , /opt/globus/setup/globus/pbs.in should be replaced with http://cyclops.phy.bg.ac.yu/mpi/pbs.in before YAIM invocation on **CE** in order to force WN to use local scratch instead of shared */home* for single CPU jobs

YAIM invocation

- **YAIM invocation command** for WN with MPI support should look like:

```
/opt/glite/yaim/bin/yaim -c -s /path/to/site-info.def -n MPI_WN -n WN -n
TORQUE_client
```

- Note that **MPI_WN** has to be first in the line

- In case that YAIM returns an error anywhere in the procedure, check data in site-info.def and other input files and restart YAIM

Tuning configured WN

- In order to adjust MPI support on newely installed WN, additional *bash* and *csh* profiles should be added. Their job is to send single-cpu jobs to local file system and multi-cpu ones to shared */home*:

```
export TMPDIR=/scratch
export EDG_WL_SCRATCH=/scratch
export PATH=/opt/edg/sbin:/opt/mpich-1.2.7p1/bin:/opt/mpiexec-0.82/bin:$PATH
if [ -n "$PBS_NODEFILE" ]; then
    if [ -r $PBS_NODEFILE ]; then
        if [ `wc -l < $PBS_NODEFILE` -gt 1 ]; then
            unset EDG_WL_SCRATCH
            unset TMPDIR
        fi
    fi
fi

setenv TMPDIR /scratch
setenv EDG_WL_SCRATCH /scratch
setenv PATH /opt/edg/sbin:/opt/mpich-1.2.7p1/bin:/opt/mpiexec-0.82/bin:$PATH
if ( $?PBS_NODEFILE ) then
 if ( -r $PBS_NODEFILE ) then
      if ( `wc -l < $PBS_NODEFILE` > 1 ) then
           unsetenv EDG_WL_SCRATCH
           unsetenv TMPDIR
      endif
 endif
endif
```

VO support

- **SEEGRID VO**
  - Install latest seegrid RPM available at http://www.irb.hr/users/vvidic/seegrid/
- **AEGIS VO**
  - Put http://voms.phy.bg.ac.yu/voms.phy.bg.ac.yu.119 into /etc/grid-security/vomsdir
- ……..

WN replication

- If all WNs are identical, there is no need to install and configure them all separately
- Following guide at http://wiki.egee-see.org/index.php/SEE-GRID_Guide_on_WN_replication, it is possible to build low-level backup of HDD into single TGZ file, and then extract the contents onto brand new HDD using any live Linux distro

```
 dd if=/dev/hda bs=1k conv=sync,noerror | gzip -c | ssh -c blowfish user@hostname "dd of=hda.gz bs=1k"

     dd if=hda.gz | ssh -c blowfish root@deadhost "gunzip -c | dd of=/dev/hda bs=1k"
```

- Afterwards, one should run **kudzu** in order to reconfigure NIC or any hardware that differs

Helpful links

- http://wiki.egee-see.org/index.php/SG_GLITE-3_Guide
- http://wiki.egee-see.org/index.php/SL4_WN_glite-3.1
- http://wiki.egee-see.org/index.php/SEE-GRID_MPI_Admin_Guide
- https://twiki.cern.ch/twiki/bin/view/EGEE/GLite31JPackage
- https://twiki.cern.ch/twiki/bin/view/LCG/YaimGuide400
- http://wiki.egee-see.org/index.php/SEE-GRID_Guide_on_WN_replication