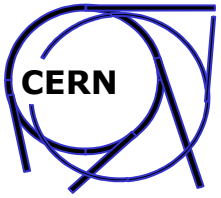


Running a LCG-2 Site

Piotr Nyczyk
CERN IT/GD

LCG-2 Administrator's Course
Oxford
19-21 July 2004

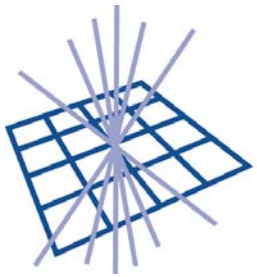
LCG2 Administrator's Course

Oxford University, 19th – 21st July 2004.

Running a LCG-2 Site

Piotr Nyczyk, CERN IT/GD

Developed in conjunction with GridPP and EGEE



GridPP
UK Computing for Particle Physics

eGee
Enabling Grids for E-science in Europe

Operations

- Changing zones: from Test Zone to Production
- Managing VOs:
  - Adding a new VO
  - Enabling/disabling VO
- Taking a site off-line:
  - for short maintenance periods
  - removing site completely from the Grid
- Batch system (PBS) management
- Maintaining correct operation of a site
- Resources to watch (Disk space and I-nodes on WN, RB, SEs)

Test and Production Zone

- Zone is simply a list of sites
  - Test Zone – new sites (under certification) tests and all other sites
  - Production Zone – certified sites, with core services running stable (each sites providing significant resources)
- Certification tests are run by Deployment Team (soon by CICs), repeated several times to check if site is stable
- Regular rerun of tests by GOC (more detailed tests soon)
- Current implementation – separate BDII servers for Test and Production, example:
  - lxn1189.cern.ch – Test Zone BDII
  - lxn1178.cern.ch – Production Zone BDII
  - Many experiment specific BDIIs managed by experiments
- Site qualification by flags in the information system
  - Lists for BDIIs still useful if managed by experiments (new tools by Oliver Keeble to ease this task).

Changing zone

- What Deplyment Team does if your site is certified and stable?
  - move your site from one list to the other (Test -> Production)
  - notify you on LCG Rollout List
- What you should do?
  - change the BDII you are using (RM configuration on WN/UI)

    In *site-cfg.h: SITE_BDII_HOST* (for LCFGng)

    In *edg-replica-manger.conf: mds.url* (manual installation)
  - change the default RB on your UI

    In *site-cfg.h: UI_RESBROKER, UI_LOGBOOK*

    Manual installation – *in /opt/edg/etc/ files:*
    - *edg_wl_ui_cmd_var.conf*
    - *<vo-name>/edg_wl_ui.conf* (for all supported VOs!)

Adding a VO

- Not an easy task with LCFGng installation – requires changes in many files
- Things that have to be set up for a new VO:
  - UI Workload Management configuration
  - grid-mapfile – mkgridmap configuration (CE/SE)
  - pool accounts and group for VO (CE/SE)
  - storage directory for VO and entries in GRIS on SE
  - common directory for VO related software (experiments software)
  - entries in information system on CE
  - additional common settings
- Example: adding BaBar VO

Adding a VO
UI Workload Management

- Resource Broker/MyProxy for UI for a new VO are defined in UserInterface-cfg.h:

```
#ifdef SE_VO_BABAR
EXTRA(uicmnconfig.vo)          babar
uicmnconfig.hlr_babar          #HLRLocation = "fake HLR Location"
uicmnconfig.myproxy_babar      MyProxyServer = MY_PROXY_SERVER
uicmnconfig.nslines_babar      babar01
uicmnconfig.nsline_babar01     UI_RESBROKER:7772
uicmnconfig.lblines_babar      babar01
uicmnconfig.lbline_babar01     UI_RESBROKER:9000
#endif
```

Adding a VO grid-mapfile

- Changes in mkgridmap-cfg.h:

```
#ifdef SE_VO_BABAR
EXTRA(mkgridmap.groups)   babarsgm
mkgridmap.uri_babarsgm    ldap://babar-vo.gridpp.ac.uk/ou=babarsgm,dc=gridpp,dc=ac,dc=uk
mkgridmap.user_babarsgm   babarsgm

EXTRA(mkgridmap.groups)   babar
mkgridmap.uri_babar       ldap://babar-vo.gridpp.ac.uk/ou=babar,dc=gridpp,dc=ac,dc=uk
mkgridmap.user_babar      .babar

EXTRA(mkgridmap.auths)    edg
mkgridmap.uri_edg         ldap://marianne.in2p3.fr/ou=People,o=testbed,dc=eu-datagrid,dc=org
#endif
```

VO can be divided into several groups

Adding a VO
Pool accounts

UNIX accounts (and groups) defined in file Users-cfg.h:

```
/* Start of babar VO */
#ifdef SE_VO_BABAR

EXTRA(auth.users)        babarsgm
auth.usercomment_babarsgm  Babar SGM Account
auth.userhome_babarsgm     /home/babarsgm
auth.usergroup_babarsgm    babar
auth.useruid_babarsgm       36352

EXTRA(auth.users)        babar001
auth.usercomment_babar012  Bfactory pool account
auth.userhome_babar012     /home/babar001
auth.usergroup_babar012    babar
auth.useruid_babar012         36000
[...]
EXTRA(auth.users)        babar050
auth.usercomment_babar012  Bfactory pool account
auth.userhome_babar012     /home/babar050
auth.usergroup_babar012    babar
auth.useruid_babar012         36049

EXTRA(auth.groups)    babar
auth.groupgid_babar    2426
```

Pool accounts additional configuration (gridmapdir entries, locks etc.) in file poolaccounts-cfg.h:

```
#ifdef SE_VO_BABAR
EXTRA(poolaccounts.usernames) \
   babar001 babar002 babar003 babar004 babar005 \
   babar006 babar007 babar008 babar009 babar010 \
   babar011 babar012 babar013 babar014 babar015 \
   babar016 babar017 babar018 babar019 babar020 \
   babar021 babar022 babar023 babar024 babar025 \
   babar026 babar027 babar028 babar029 babar030 \
   babar031 babar032 babar033 babar034 babar035 \
   babar036 babar037 babar038 babar039 babar040 \
   babar041 babar042 babar043 babar044 babar045 \
   babar046 babar047 babar048 babar049 babar050
#endif
```

Adding a VO
Storage directory/GRIS on SE

- Storage directory (flatfiles) defined in flatfiles-dirs-SECLASSIC-cfg.h:

```
#ifdef SE_VO_BABAR
EXTRA(dirperm.ents)          babar
dirperm.path_babar           CE_CLOSE_SE_MOUNTPOINT/SA_PATH_BABAR
dirperm.owner_babar          root:babar
dirperm.perm_babar           0775
dirperm.type_babar           d
#endif
```

- Entries in GRIS on SE defined in file lcginfo-seclassic-cfg.h:

```
#ifdef SE_VO_BABAR
#define PATH_BABAR babar:SA_PATH_BABAR
#define PATH_DYN_BABAR babar:CE_CLOSE_SE_MOUNTPOINT/SA_PATH_BABAR

EXTRA(lcginfo.args_classic) PATH_DYN_BABAR

EXTRA(lcginfo.entry) babarSA
#define DN_BABARSA dn: GlueSARoot=PATH_BABAR,GlueSEUniqueID=SE_HOSTNAME,Mds-Vo-name=local,o=grid
lcginfo.dn_babarSA DN_BABARSA
lcginfo.attributes_babarSA GlueChunkKey GlueSAAccessControlBaseRule
lcginfo.values_babarSA_GlueChunkKey GlueSEUniqueID=SE_HOSTNAME
lcginfo.values_babarSA_GlueSAAccessControlBaseRule babar
#endif
```

Adding a VO
Directory for VO related software

- Common directory for VO related software:
  - environmental variable for WN is defined in WorkerNode-cfg.h:

```
#ifdef SE_VO_BABAR
EXTRA(lcgenv.name)                babar
lcgenv.variable_babar             VO_BABAR_SW_DIR
lcgenv.value_babar                WN_AREA_BABAR
#endif
```

Note! WN_AREA_BABAR variable is defined in site-cfg.h (next slide) and the corresponding directory should be set up in a site specific way

  - the directory for VO software manager to publish information about installed software is defined in file voswmgr-dirs-cfg.h:

```
#ifdef SE_VO_BABAR
EXTRA(dirperm.ents) babarsgm
dirperm.path_babarsgm                INFO_PATH/babar
dirperm.owner_babarsgm               babarsgm:babar
dirperm.perm_babarsgm                0755
dirperm.type_babarsgm                d
#endif
```

Adding a VO additional settings

- Define VO as supported on CE in file ComputingElement-novoms-cfg.h

```
#ifdef SE_VO_BABAR
EXTRA(ceinfo.SUPPORTEDVOS) babar
#endif
```

- Add *edguser* to the group of the VO in user-edguser-cfg.h file:

```
#ifdef SE_VO_BABAR
EXTRA(auth.usersuppgroups_edguser) babar
#endif
```

- Definitions in site-cfg.h:

```
#define SE_VO_BABAR
#define SA_PATH_BABAR babar
#define WN_AREA_BABAR   /opt/exp_software/babar # see previous slide
```

- After changing the files and recompiling profiles make sure that all changes were propagated to all machines (CE, SE, WNs, UI). The easiest way is to reboot all machines.

Disabling/Enabling VO

- Why? To temporarily stop attracting new jobs for one or more VOs
- Edit */opt/edg/var/etc/ce-static.ldif* file on your CE and remove *GlueCEAccessControlBaseRule* entries for the VO in all execution queues.
- Restart *globus-mds* service on CE.
- No jobs will be killed, site will only stop to attract new jobs for particular VO.
- To check if site is not publishing particular VO as supported try to use *edg-job-list-match* command.
- To prevent jobs from a VO running on your site edit the configuration file for mkgridmap and remove line(s) related to the particular VO
- To enable the VO put the entry in *ce-static.ldif* file back and restart *globus-mds* service. Restore the contents of mkgridmap configuration file.

Taking a site off line

- Non trivial task (especially when removing site completely from grid)
- First – disable all VOs to stop accepting any new jobs (previous slide – edit mkgridmap configuration file)
- Watch PBS queues if there are still jobs running or waiting in queues – wait for all jobs to finish
- Do not stop any services before all jobs finish!
- When all jobs are done – you can safely stop services/machine(s) for short maintenance period
- For long time shut down (or complete removal) of a site additional steps are required

Removing a site from Grid

- site off-line - all jobs finished cleanly, BUT data is still there!
- If your site is operating RB:
  - stop accepting new jobs on your RB (HOW???)
  - wait 10 days to expire sandboxes
- Contact VO managers to remove data from your SE
- The procedure is quite long and it depends on others (VO managers...)

Basic PBS Operations

- *qstat* – checking the status of queues (options: -q -n ...)
- *pbsnodes* – checking nodes status and simple node manipulation (options -a -l -c ...)
- *qhold / qrls* – hold / release a job
- *qrun* – force execution of a job regardless of scheduling position
- *qdel* – terminate (TERM, KILL) a job and remove it from batch system
- *qmove* – move a job to different queue
- *qmgr* – PBS batch system manager:
  - set server/scheduler parameters
  - define/modify queues and nodes

Maintaining correct operation of a site

- Test your site from time to time
- Read carefully messages on roll-out list, especially Test Zone reports
- Check if your batch system works correctly: *qstat* command should return without delays
- Check if your GIIS is working and if it is providing correct (consistent) information (ldapsearch, ldap browser...)
- Run certification test script
- In case of problems try to reproduce them (commands from script) and analyse output
- Watch disk space on RB, WNs and SEs (have a look at swap on WNs)