

# DPM Administration for Tier2s

*Sophie Lemaitre* ([Sophie.Lemaitre@cern.ch](mailto:Sophie.Lemaitre@cern.ch))

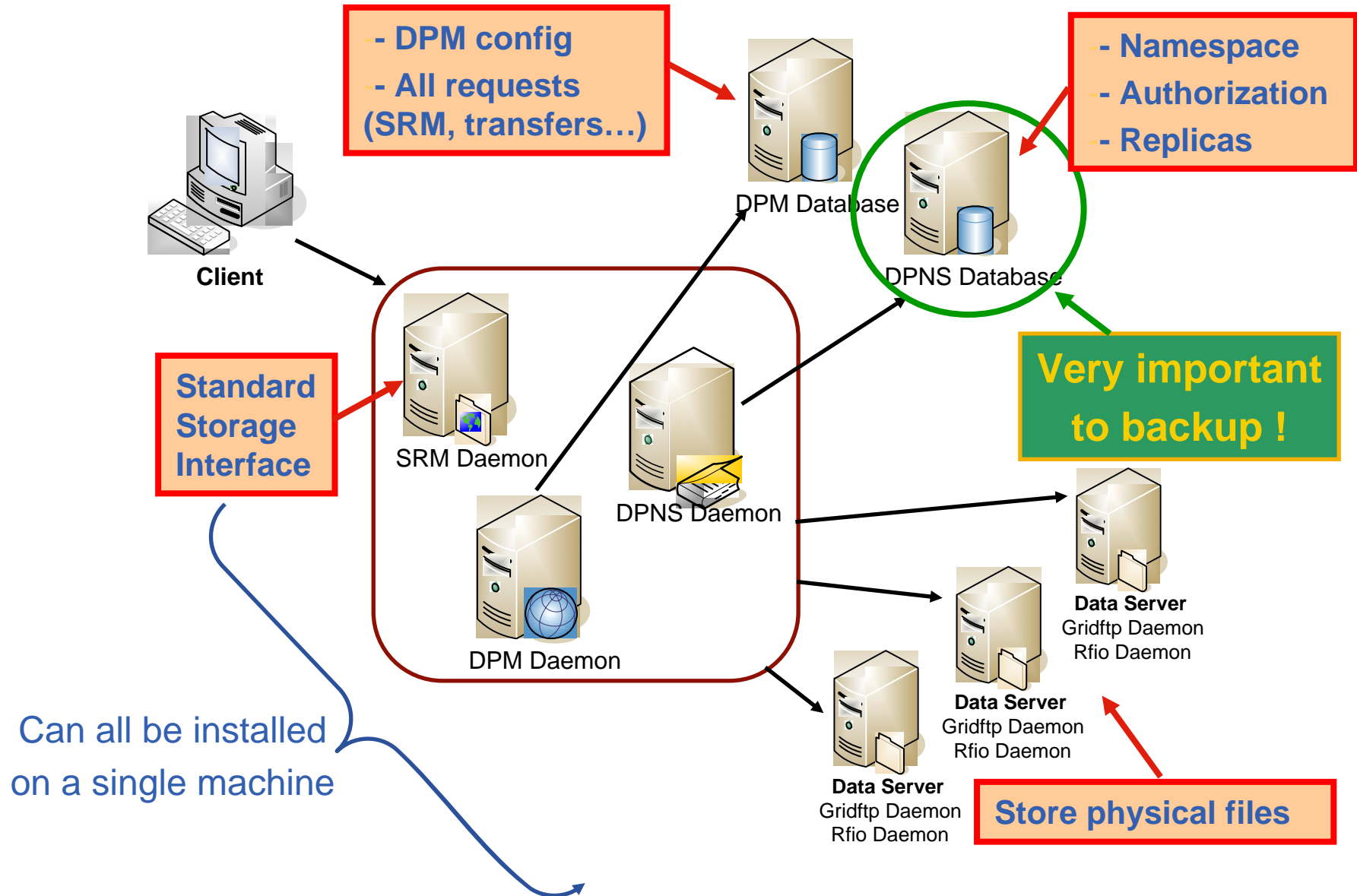
*Jean-Philippe Baud* ([Jean-Philippe.Baud@cern.ch](mailto:Jean-Philippe.Baud@cern.ch))

*Asia Tier2s tutorial – 2<sup>nd</sup> Dec 2006*

- **Description**
- **Installation**
- **DPM as a service**
- **Troubleshooting**
- **Documentation / support**

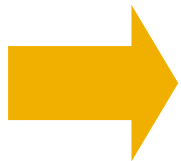
- **Description**
- **Installation**
- **DPM as a service**
- **Troubleshooting**
- **Documentation / support**

- **Disk Pool Manager**
  - Manages storage on disk servers
  - SRM support
    - 1.1
    - 2.1
    - 2.2 coming soon
- **Deployment status**
  - 83 DPMs in production
  - 110 VOs supported



- **Easy to install/configure**
  - **Few configuration files**
- **Manageable storage**
  - **Logical Namespace**
  - **Easy to add/remove file systems**
- **Low maintenance effort**
- **Supports as many disk servers as needed**
- **Low memory footprint**
- **Low CPU utilization**
- **Easy Classic SE to DPM migration**

- DPM server
- DPM Name Server
- SRM servers
  - srm v1
  - srm v2
- RFIO server
- DPM-enabled GridFTP server
  - Normal GridFTP server slightly modified for the DPM
  - Can be used in place of the normal GridFTP server
  - **But** normal GridFTP server cannot be used for the DPM
  - Note: GridFTP2 support coming soon



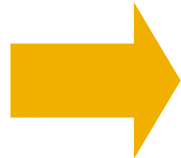
- **Description**
- **Installation**
- **DPM as a service**
- **Troubleshooting**
- **Documentation / support**



- 5 questions before installation:

- For each VO, what is the expected load?

- Does the DPM need to be installed on a separate machine ?



**Yes, this is recommended !**

- How many disk servers do I need ?

- Disk servers can easily be added or removed later

- Which operating system ?

- Which file system type ?

- At my site, can I open ports:

- 5010 (Name Server)
- 5015 (DPM server)
- 8443 (srmv1)
- 8444 (srmv2)
- 5001 (rfio)
- 20000-25000 (rfio data port)
- 2811 (DPM GridFTP control port)
- 20000-25000 (DPM GridFTP data port)

- **RPMs released for:**
  - **SLC3**
  - **Itanium**
  
- **DPM builds on:**
  - **Mac OS X**
  - **SLC4**
  - **Other 64 bit platforms**

- Dedicated file systems
  - DPM cannot work properly if each file systems is not dedicated:
    - Reported space will be incorrect
    - Thus, space reservation, garbage collection, etc. will not work as expected



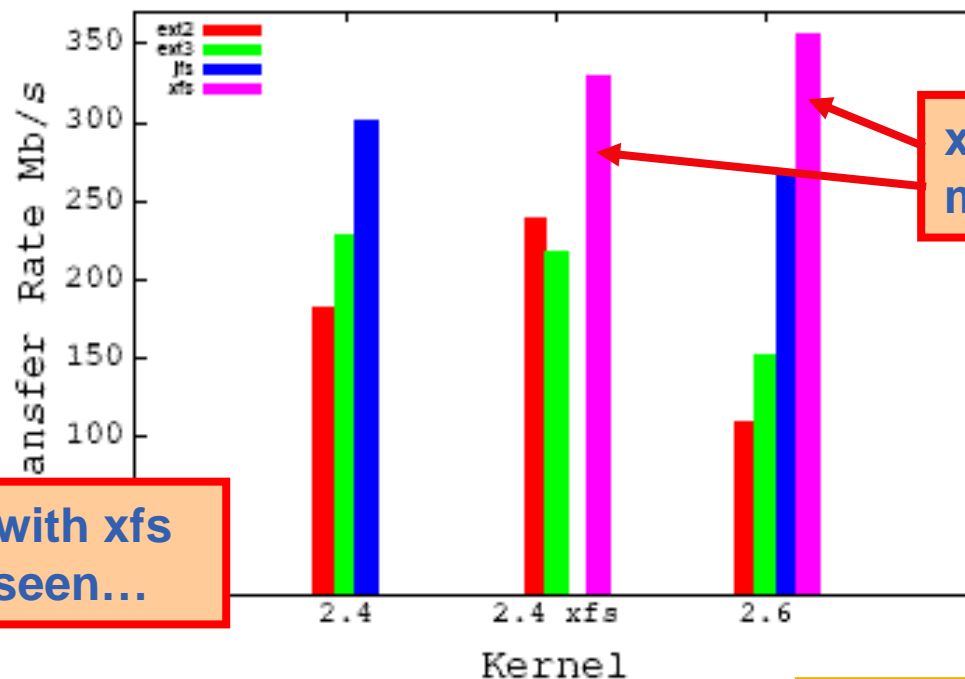
**VERY IMPORTANT**

- Supported file systems ?
  - ext2: not recommend
  - ext3, xfs: OK
  - reiserfs: ?
- NFS mounted file system ?
  - Problems seen:
    - DPM sometimes hangs
    - File corruption

# Which File System type ?



DPM:  $N_f = N_s = 5$



xfs seems to be more performant

But, problems with xfs and kernel 2.6 seen...

Small # failed file transfers for:

- ext2,3 with 2.6 kernel and jfs with vanilla 2.4 kernel.

From Greig A. Cowan (HEPIX 2006)

- **YAIM configuration :**

**Site-info.def**

\$MY_DOMAIN	\$DPM_DB_PASSWORD
“\$DPM_HOST:/data01 disk1.\$MY_DOMAIN:/data02”	B_HOS
\$DPM_HOST	\$DPMFSIZE
\$DPMPOOL	\$SE_LIST
\$DPM_FILESYSTEMS	\$SE_ARCH
\$DPM_DB_USER	\$VOS

Default amount of space reserved for a file. Check VO typical file usage.

“multidisk”

- VO\_<myVO>\_DEFAULT not used for DPM
- VO\_<myVO>\_STORAGE\_DIR not used for DPM

- **On the DPM server:**

- ./install\_node site-info.def glite-SE\_dpm\_mysql
- ./configure\_node site-info.def glite-SE\_dpm\_mysql

- **On each disk server (except DPM server):**

- ./install\_node site-info.def glite-SE\_dpm\_disk
- ./configure\_node site-info.def glite-SE\_dpm\_disk

- **YAIM installs**
  - **On \$DPM\_HOST:**
    - DPM server
    - DPM Name Server
    - SRM servers (srmv1 & srmv2)
  - **On each disk server specified in \$DPM\_FILESYSTEMS:**
    - RFIO server
    - DPM-enabled GridFTP server
  
- **YAIM creates**
  - **One pool including all file systems specified**
  
- **Only configuration files**
  - `/opt/lcg/etc/DPMCONFIG` : DPNS and DPM DB connection string
  - `/etc/shift.conf` : different servers/disk servers TRUST rules

- **RFIO, GridFTP run as root**
- **Dedicated user/group**
  - DPM, DPNS, SRM daemons run as *dpmmgr*
  - Several directories/files belong to *dpmmgr*
- **Host certificate, key**

```

> ll /etc/grid-security/ | grep pem
-rw-r--r--  1 root  root    5430 May 28 22:02 hostcert.pem
-r-----  1 root  root    1675 May 28 22:02 hostkey.pem

> ll /etc/grid-security/dpmmgr/ | grep pem
-rw-r--r--  1 dpmmgr dpmmgr  5430 May 28 22:02 dpmcert.pem
-r-----  1 dpmmgr dpmmgr  1675 May 28 22:02 dpmkey.pem

```

- **Database connect**
  - /opt/lcg/etc/NSCONFIG
  - /opt/lcg/etc/DPMCONFIG
    - <username>/<password>@<mysql\_server>
- **Sysconfig files**
  - /etc/sysconfig/<service\_name>
    - Number of threads (only DPNS)
- **Daemons**
  - service <service\_name> {start|stop|status}
  - Important: services not restarted by RPM upgrade !



# How does it look ?

- DPM Pool**

```
export DPM_HOST=dpm01.cern.ch
dpm-qryconf
```

Cleanup starts when 5% space left

```
POOL TutorialPool DEFSIZE 200.00M GC_START_THRESH 5
GC_STOP_THRESH 15 DEFPINTIME 0 PUT_RETENP 86400
FSS_POLICY maxfreespace GC_POLICY lru RS_POLICY fifo GID 0
CAPACITY 52.77G FREE 43.79G ( 83.0%)
data CAPACITY 17.27G FREE 14.24G ( 82.4%)
disk01.cern.ch /storage CAPACITY 17.75G FREE 14.78G ( 83.3%)
disk01.cern.ch /data CAPACITY 17.75G FREE 14.78G ( 83.3%)
```

Cleanup stops when 15% space left

- Name Server**

```
export DPNS_HOST=dpm01.cern.ch
dpns-ls -l /dpm/cern.ch/home
```

Your domain

Supported VOs

```
drwxrwxr-x 0 root 2689 0 May 22 14:43 atlas
drwxrwxr-x 2 root 2688 0 May 22 14:53 dteam
Virtual gids drwxrwxr-x 0 root 2690 0 May 22 14:43 geant4
drwxrwxr-x 0 root 2691 0 May 22 14:43 lhcb
```

Virtual gids

- **lcg\_utils (from a UI)**
  - **If DPM not in site BDII yet**
    - `export LCG_GFAL_INFOSYS=dpm01.cern.ch:2135`
    - `lcg-cr -v --vo dteam -d dpm01.cern.ch file:/path/to/file`
  - **Otherwise**
    - `export LCG_GFAL_INFOSYS=my_bdii.cern.ch:2170`
    - `lcg-cr -v --vo dteam -d dpm01.cern.ch file:/path/to/file`
  
- **rfio (from a UI)**
  - `export LCG_RFIO_TYPE=dpm`
  - `export DPNS_HOST=dpm01.cern.ch`
  - `export DPM_HOST=dpm01.cern.ch`
  - `rmdir /dpm/cern.ch/home/dteam/myVO`
  
- **Note: common RFIO library coming in next release, no need to set \$LCG\_RFIO\_TYPE then.**

- **srmcp (from a UI)**
  - Works with DPM as source
  - Doesn't work yet if DPM as target
    - /opt/d-cache/srm/bin/srmcp file:///path/to/file  
srm://dpm01.cern.ch:8443/ **SRM server** me/dteam/my\_dir/my\_file
  
- **globus-url-copy (from a UI)**
  - lcg-gt srm://dpm01.cern.ch/dpm/cern.ch/home/dteam/generated/2006-06-14/file5b517db2-30dc-4df0-9ac9-0e30433e10da gsiftp

TURL returned:

  - globus-url-copy file:/path/to/file  
gsiftp://disk01.cern.ch/disk01.cern.ch:/data/dteam/2006-06-14/file5b517db2-30dc-4df0-9ac9-0e30433e10da.1.0
  - globus-url-copy file:/path/to/file  
gsiftp://disk01.cern.ch/disk01.cern.ch:/data/dteam/2006-06-14/file5b517db2-30dc-4df0-9ac9-0e30433e10da.1.0

- **Description**
- **Installation**
- **DPM as a service**
  - Administration
  - Monitoring
- **Troubleshooting**
- **Documentation / support**

- **Why migrating your Classic SE ?**
  - **Users:**
    - SRM support
    - Logical Namespace
  - **Admins:**
    - Manageable storage (easy to add/remove disk space)
    - Automatic garbage collection
- **Migrating to what ?**
  - **DPM**
    - Only metadata operation (Name Server population)
    - No physical file move needed
- **Script to run**

```
./migration classicSE_hostname classicSE_dir  
DPM_hostname DPM_dir DPM_pool
```
- **More details, see**  
<https://twiki.cern.ch/twiki/bin/view/LCG/ClassicSeToDpm>

- **Modify configuration**

```
dpm-modifypool --poolname myPool --gc_start_thresh 5 --
gc_stop_thresh 10
```

```
dpm-modifyfs --server disk01.cern.ch --fs /data --st RDONLY
```

- **Add a disk server**

- `./install_node site-info.def glite-SE_dpm_disk`

- `./configure_node site-info.def glite-SE_dpm_disk`

- **Add it to TRUST rules in DPM server** `/etc/shift.conf` **file**

- **Remove a disk server**

- `dpm-drain --server disk02.cern.ch --fs /data`

- `dpm-rmfs --server disk02.cern.ch --fs /data`

- **Remove it from TRUST rules in DPM server** `/etc/shift.conf` **file**

- **Load balancing**
  - **DPM automatically round robins between file systems**
  
- **Example**
  - **disk01: 1TB file system**
  - **disk02: very fast, 5TB file system**
  
  - **Solution 1:       one file system per disk server**
    - A file will be stored on either disk, equally, if space left
  - **Solution 2:       one file system on disk01**  
**two file systems on disk02**
    - A file will more often end up on disk02, which is what you want

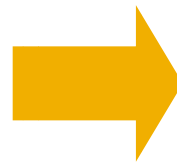
- **dpm-replicate**

- Users and admins can do it
- Useful if file often accessed

```
dpm-replicate -s_type P /dpm/cern.ch/home/dteam/my_file
```

- **dpm-drain**

- Sets file system(s) to READ-ONLY
- Replicates to another pool/file system
  - Pinned files
  - Permanent files



**Empty pool / file system  
can then be removed physically**

```
dpm-drain --poolname myPool
```

```
dpm-drain --server disk01.cern.ch
```

```
dpm-drain --server disk01.cern.ch --fs /data
```



- **Remove Name Server entries**
  - **Use rfrm (as user or root)**
    - `export DPNS_HOST=dpm01.cern.ch`
    - `export DPM_HOST=dpm01.cern.ch`
    - `rfrm /dpm/cern.ch/home/dteam/my_dir/my_file`
  - **Removes**
    - Files on disk
    - Replicas and Logical Files Names in DPM Name Server

- **Dedicate a Pool to a VO**

- **By default, pool can be used by all supported groups/VOs**

- **But, pool can be restricted:**

- `dpm-modifypool --poolname VOpool --def_filesize 200M --gid VO_gid`

- `dpm-modifypool --poolname VOpool --def_filesize 200M --group VO_group_name`

- **Hard limit:**

- What to do if pool is full ?

- **Quotas**

- **Maximum space allowed per group/VO**

- **Softer limit:**

- Not bind to pools / physical file systems

**Not implemented yet**

- **Dynamic space reservation**

- **Defined by user on request**

- **Maximum value configurable**

- **Admin maximum value could be higher than for standard user**

**Not released yet**

- **What to monitor ?**
  - All processes running ?
  - All threads busy ? **DPNS**
  - All threads busy ? **DPM**
    - 20 threads for fast operations
    - 20 threads for slow operations (get, put, copy)
    - Note: there are also
      - 1 garbage collector thread per disk pool defined
      - 1 thread that removes the expired put requests
  - **RFIO/GridFTP**
    - Too many transfers to the same file system ?
  - **Number of requests (in `dpm_db`) per week/month ?**

Monitor separately

- **GridView**

- **Monitoring tool for GridFTP transfers**
- **On each disk server, enable GridView**
  - Edit `/opt/lcg/etc/lcg-mon-gridftp.conf`
  - Change: `LOG_FILE = /var/log/dpm-gsiftp/dpm-gsiftp.log`
  - Restart service: `service lcg-mon-gridftp restart`

Not supported by YAIM yet

- **Request DB cleanup script**
  - **Request DB only grows:**
    - Ex: two more rows in the `dpm_db` database, when opening a file
  - **Need script to remove the older requests**
  - **Admins can configure how long the requests are kept**
- **Implement quotas**
- **Deploy dynamic space reservation**
- **Limitation of transfers to same file system**
- **Automated database backups for MySQL**
  - **MySQL replication**
  - **Save/backup data from the slave in a safe place (Tier1 ?)**
  - **Merge**

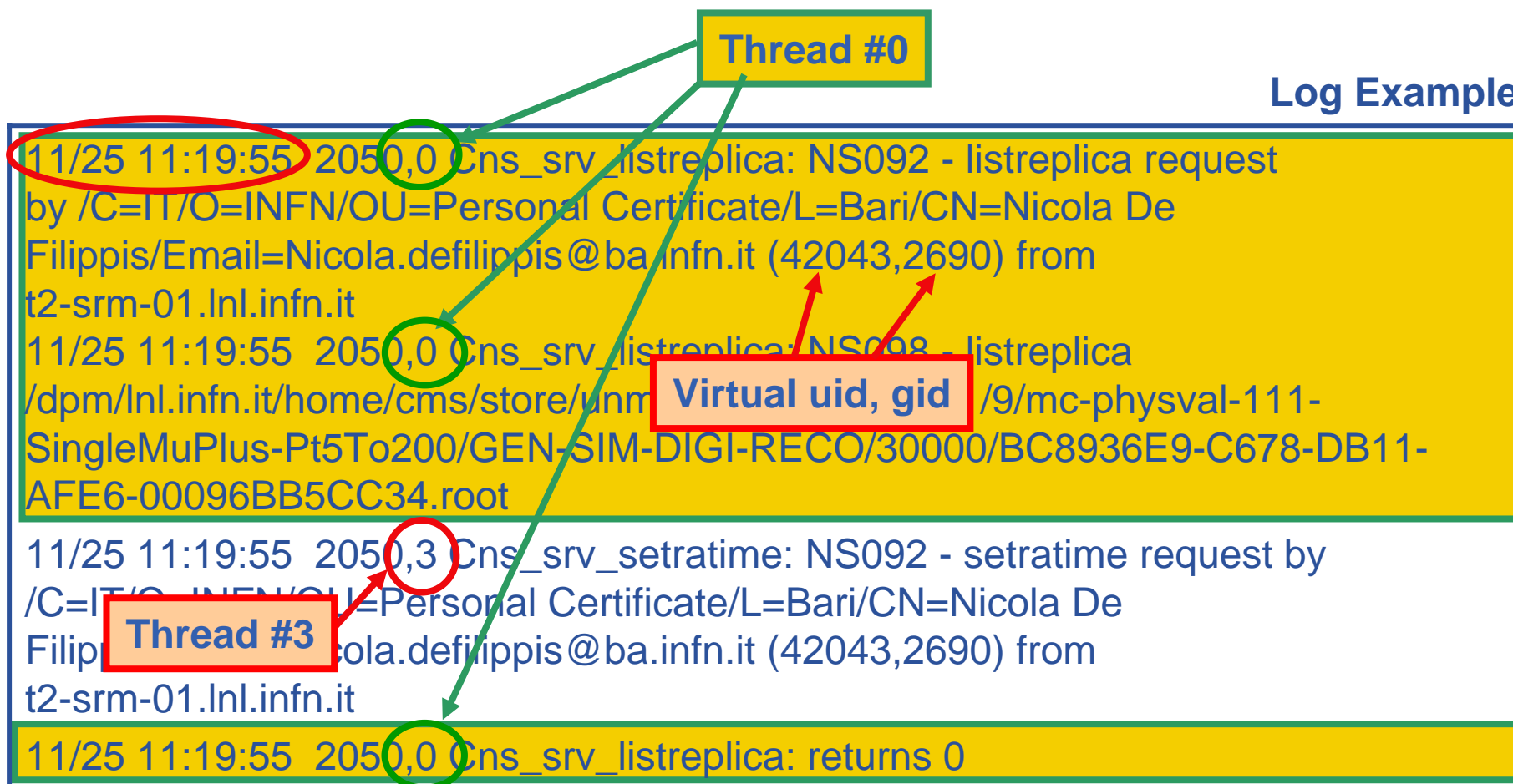
- **Description**
- **Installation**
- **DPM as a service**
- **Troubleshooting**
- **Documentation / support**

- **DPM server**
  - `/var/log/dpm/log`
- **DPM Name Server**
  - `/var/log/dpns/log`
- **SRM servers**
  - `/var/log/srmv1/log`
  - `/var/log/srmv2/log`
- **RFIO server**
  - `/var/log/rfiod/log`
- **DPM-enabled GridFTP**
  - `/var/log/dpm-gsiftp/gsiftp.log`
  - `/var/log/dpm-gsiftp/dpm-gsiftp.log`

- **Example:**

- /var/log/dpns/log

## Log Example



The diagram illustrates a sequence of log entries from a DPM log file. A yellow box highlights the first three entries, which are associated with Thread #0. A red box highlights the fourth entry, associated with Thread #3. A red box labeled 'Virtual uid, gid' points to the '2050,0' and '2050,3' values in the log entries. Green circles and arrows indicate the thread associations: Thread #0 for the first three entries and Thread #3 for the fourth entry.

```

11/25 11:19:55 2050,0 Cns_srv_listreplica: NS092 - listreplica request
by /C=IT/O=INFN/OU=Personal Certificate/L=Bari/CN=Nicola De
Filippis/Email=Nicola.defilippis@ba.infn.it (42043,2690) from
t2-srm-01.lnl.infn.it
11/25 11:19:55 2050,0 Cns_srv_listreplica: NS098 - listreplica
/dpm/lnl.infn.it/home/cms/store/unm /9/mc-physval-111-
SingleMuPlus-Pt5To200/GEN-SIM-DIGI-RECO/30000/BC8936E9-C678-DB11-
AFE6-00096BB5CC34.root
11/25 11:19:55 2050,3 Cns_srv_setratime: NS092 - setratime request by
/C=IT/O=INFN/OU=Personal Certificate/L=Bari/CN=Nicola De
Filippis/Email=Nicola.defilippis@ba.infn.it (42043,2690) from
t2-srm-01.lnl.infn.it
11/25 11:19:55 2050,0 Cns_srv_listreplica: returns 0
    
```

- Rotated logs kept for 90 days

- > ls -al /var/log/srmv1/

total 27420

```
drwxr-xr-x  2 dpmmgr  dpmmgr  4096   Jul 25 04:02  .
drwxr-xr-x 13 root    root    4096   Jul 25 00:20  ..
-rw-r--r--  1 dpmmgr  dpmmgr 906619 Jul 25 15:42  log
-rw-r--r--  1 dpmmgr  dpmmgr 3582741 Jul 25 04:01  log.1
-rw-r--r--  1 dpmmgr  dpmmgr  71944  Jul 16 04:00  log.10.gz
-rw-r--r--  1 dpmmgr  dpmmgr 121344  Jul 15 04:01  log.11.gz
-rw-r--r--  1 dpmmgr  dpmmgr 388301  Jul 14 04:01  log.12.gz
...
-rw-r--r--  1 dpmmgr  dpmmgr 140740  Jul  7  04:00  log.19.gz
-rw-r--r--  1 dpmmgr  dpmmgr  32601  Jul 24 04:01  log.2.gz
-rw-r--r--  1 dpmmgr  dpmmgr  92150  Jul  6  04:01  log.20.gz
-rw-r--r--  1 dpmmgr  dpmmgr  77918  Jul  5  04:01  log.21.gz
-rw-r--r--  1 dpmmgr  dpmmgr  15468  Jul  4  03:59  log.22.gz
...
```



- **Security**
  - Host certificate, key
  - CRLs
- **VOMS setup**
  - On LFC & UI, `/etc/grid-security/vomsdir` contains VO VOMS server

```
> ls -ld /etc/grid-security/vomsdir/
drwxr-xr-x  2 root  root 4096 Jun  8 15:07 /etc/grid-security/vomsdir/
```

```
> ls /etc/grid-security/vomsdir
cclcgvomsli01.in2p3.fr.43
lcg-voms.cern.ch.1265
```

- On the UI (client), `/opt/glite/etc/vomses`

```
> ls /opt/glite/etc/vomses
alice-lcg-voms.cern.ch
alice-voms.cern.ch
```

- **dpns-ls -l**

**Virtual uid, gid**

```
> dpns-ls -l /dpm/cern.ch/home/dteam/subdirx
-rwxr-xr-- 1 18964 2688 6647 Jul 18 10:57 srmv1_Tfile.txtA0
-rw-rw-r-- 1 18964 2688 129 Jul 18 10:59 srmv1_Tfile.txtA1
```

- **dpns-getacl**

**Actual DN, group**

```
> dpns-getacl /dpm/cern.ch/home/dteam/subdirx
# file: /dpm/cern.ch/home/dteam/subdirx
# owner: /C=CH/O=CERN/OU=GRID/CN=Martin Flechl 5835
# group: dteam
user::rwx
group::r-x          #effective:r-x
other::r-x
default:user::rwx
default:group::rwx
default:other::r-x
```

- **User DN**

- Added automatically to `Cns_userinfo`, if it doesn't exist

**Cns\_userinfo**

UID	Name
18964	/C=CH/O=CERN/OU=GRID/CN=Martin Flechl 5835
18967	/C=CH/O=CERN/OU=GRID/CN=Sophie Lemaitre 2268

- **User group**

- grid-proxy-init or simple voms-proxy-init
  - Group taken from `/opt/lcg/etc/lcgdm-mapfile`

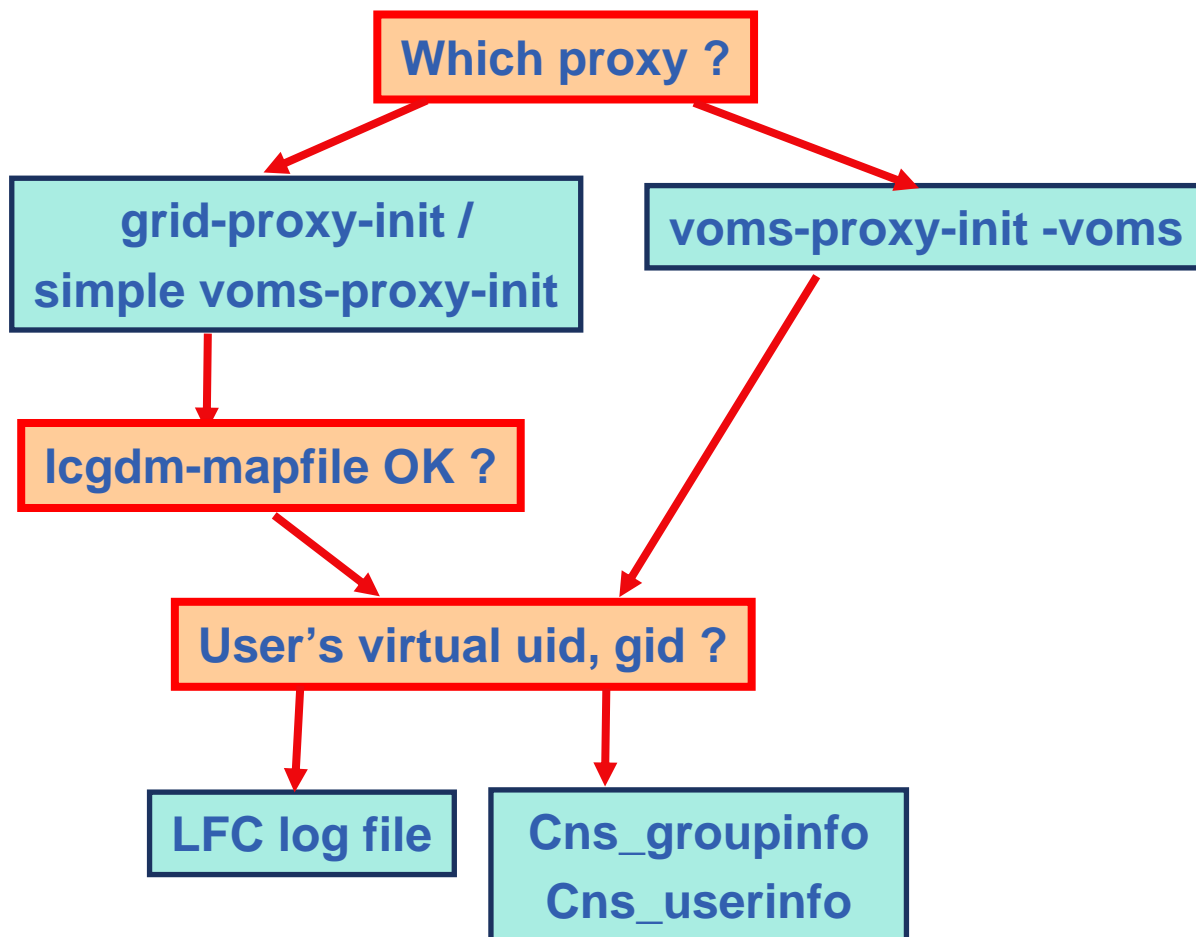
"/C=CH/O=CERN/OU=GRID/CN=Simone Campana 7461 - ATLAS" atlas  
 "/C=CH/O=CERN/OU=GRID/CN=Sophie Lemaitre 2268" dteam

- voms-proxy-init –voms myVO
  - Group taken from the VOMS role
  - Added automatically, if doesn't exist

**Cns\_groupinfo**

GID	Name
2688	dteam
2694	dteam/Role=lcgadmin

- Check user's mapping:



- **Description**
- **Installation**
- **DPM as a service**
- **Troubleshooting**
- **Documentation / support**

- **Main DPM documentation page**
  - <https://twiki.cern.ch/twiki/bin/view/LCG/DataManagementTop>
- **DPM Admin Guide**
  - <https://twiki.cern.ch/twiki/bin/view/LCG/DpmAdminGuide>
- **Troubleshooting page**
  - <https://twiki.cern.ch/twiki/bin/view/LCG/LfcTroubleshooting>

- **Contact GGUS [helpdesk@ggus.org](mailto:helpdesk@ggus.org)**
  - your ROC will help
  - If needed, DPM experts will help

## Questions ?

[helpdesk@ggus.org](mailto:helpdesk@ggus.org)

**Sophie Lemaitre** ([Sophie.Lemaitre@cern.ch](mailto:Sophie.Lemaitre@cern.ch))

**Jean-Philippe Baud** ([Jean-Philippe.Baud@cern.ch](mailto:Jean-Philippe.Baud@cern.ch))



