How to Install and Use the DQ2 User Tools

US ATLAS Tier2 workshop at IU
June 20, 2007 - Bloomington, IN

Marco Mambelli
University of Chicago

Topics

- No grid-setup
- No certificate/proxy

- DQ2
- Move data between local SE, Tier1/2, Tier3 and local disks
- Install/configure user tools
- Use user tools

Some technology

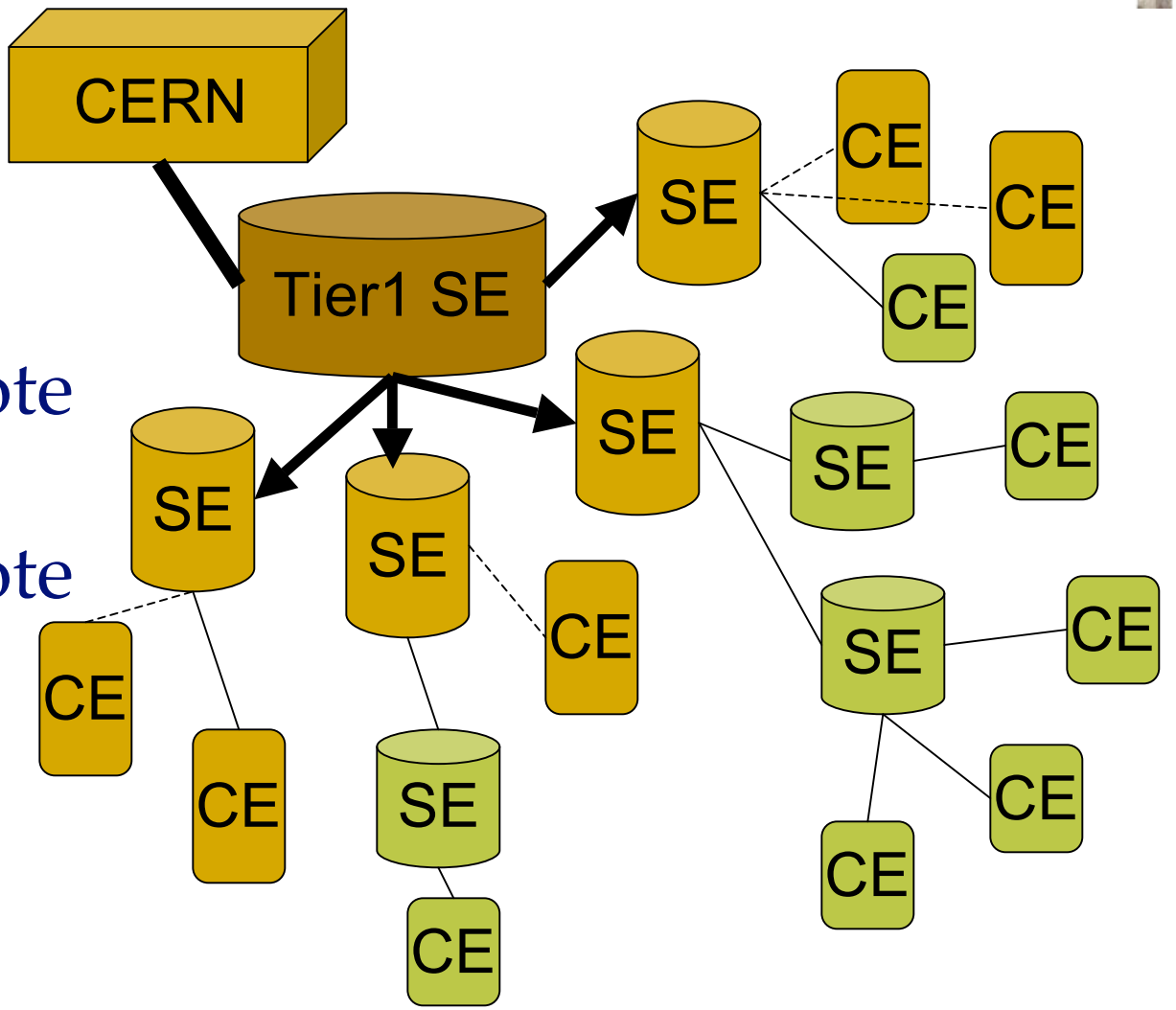
- 3 Grids in ATLAS (OSG, LCG, NG):
  - Different file catalogs:
    - Local Replica Catalog (LRC) - is used in OSG
    - Local File Catalog (LFC) - is used in LCG
  - Different information system
  - Different environment
- Software
  - Mostly python
    - Platform independent
    - This would be OK if only OSG/NG
  - LCG UI
    - Os dependent
    - Necessary to interact with LCG

DDM Overview

- DQ2 has four main components:
  - Central Dataset Catalogues (CERN)
    - responsible for bookkeeping information
  - Site services (Tier1/2)
    - responsible for fulfilling data movement requests (subscriptions)
  - Client tools
    - to interact with central dataset catalogues and site services.
  - The LCG/ARDA Dashboard
    - for monitoring data movement requests
- https://twiki.cern.ch/twiki/bin/view/Atlas/DistributedDataManagement
- http://www.usatlas.bnl.gov/twiki/bin/view/Admins/DQ2SiteServicesP1

Topology example

- Tier0/1
- Tier2/3
- FTS
- Local/remote connection
- Local/remote DQ2 (different transfer protocols)



Site services

- Subscriptions work only between sites with DQ2 Site Services
- DQ2 Will be available at Tier1 and Tier2s
  - Patrick (Friday)
- Tier3 may install it
- Tier3 can use the DQ2 service of a 'neighboring' Tier2
  - Local files not in the catalog (2 step transfers)
  - Integrated catalog and SE (LRC required)
  - Integrated SE (donate their SE to the Tier2)
- http://dashb-build.cern.ch/ddm/build/stable/doc/guides/dq2-siteservices/html/admin/

DQ2 Clients

- Necessary to access DQ2 data
- Available for local installation or through AFS
- DQ2 (Catalog) Client
  - Allows to manage subscriptions (transfer files between DQ2 sites)
  - https://twiki.cern.ch/twiki/bin/view/Atlas/DDMClientDQ2
- DQ2 User Tools
  - Allows to transfer files and datasets to local/remote directories
  - https://twiki.cern.ch/twiki/bin/view/Atlas/UsingDQ2
- Data Moving Utilities (DMU), used by Panda and DDS

Catalog Clients

- Use from AFS (For latest production release - 0.3.x):
  - Setup:

```
/afs/cern.ch/atlas/offline/external/GRID/ddm/pro03/dq2.sh
/afs/usatlas.bnl.gov/Grid/Don-Quijote/DQ2_0_3_client/setup.sh
```

- Installation (0.2 ?):
  - Probably just install a copy of the directory tree and fix the setup file

http://atlas.web.cern.ch/Atlas/GROUPS/DATABASE/project/ddm/releases/

dq2-register-subscription -a users.MarcoMambelli.testdataset1 UC

User Tools

- Access through AFS (source configuration file):
  - Setup LCG UI (or a grid environment)
    - NameError: global name 'lfc' is not defined
      The lfc module is required to access LCG datasets. It is available in LCG UI or RPM. Then, lfc.py needs to be in one of the directories in PYTHONPATH.
      https://twiki.cern.ch/twiki/bin/view/LCG/TarUIInstall
      LFC-interfaces (SL3, SL4 as well):
      http://glitesoft.cern.ch/EGEE/gLite/APT/R3.0/rhel30/RPMS.Release3.0/
  - Different setup: setup.[c,z]sh.[CERN,BNL,any]: e.g. setup.sh.any

```
/afs/cern.ch/atlas/offline/external/GRID/ddm/endusers/
/afs/usatlas.bnl.gov/Grid/Don-Quijote/dq2_user_client/
```

- Download from the CVS repository (Web accesible):
  - http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/DataManagement/DQ2_0_2/endusers/?only_with_tag=MAIN
- Doc: https://twiki.cern.ch/twiki/bin/view/Atlas/UsingDQ2

User Tools Setup

- **DQ2_URL_SERVER,** DQ2 server
  ```
  $ export DQ2_URL_SERVER=http://atlddmpro.cern.ch:8000/dq2/
  ```
- **DQ2_URL_SERVER_SSL,** DQ2 server for secure connection
  ```
  $ export DQ2_URL_SERVER_SSL=https://atlddmpro.cern.ch:8443/dq2/
  ```
- **DQ2_LOCAL_ID,** local DQ2 site ID (from TiersOfATLAS)
  Depends on where you (e.g. CERN/lxplus, BNL users)
  ```
  $ export DQ2_LOCAL_ID=CERN
  $ export DQ2_LOCAL_ID=BNL
  ```
  - For other site users, ask your DDM admin for the ID. If your site doesn't deploy a DQ2 site service, leave it blank

  ```
  $ export DQ2_LOCAL_ID=
  ```
- **DQ2_LOCAL_PROTOCOL,** Protocol to access the local storage (rfio,castor,dcap,unix,dpm). (e.g. For CERN/lxplus, BNL users)
  ```
  $ export DQ2_LOCAL_PROTOCOL=castor
  $ export DQ2_LOCAL_PROTOCOL=dcap
  ```
  - If you use normal disk storage:

  ```
  $ export DQ2_LOCAL_PROTOCOL=unix
  ```

User Tool Setup (cont)

- **DQ2_STORAGE_ROOT,** root directory of local storage
  `$ export DQ2_STORAGE_ROOT=/pfns`
- **DQ2_SRM_HOST,** local SRM server (if any)
  `$ export DQ2_SRM_HOST=srm://castorgrid.cern.ch:8443`
- **DQ2_GSIFTP_HOST,** local GridFTP server (if any)`$ export DQ2_GSIFTP_HOST=gsiftp://castorgrid.cern.ch:2811`
- **DQ2_USE_SRM,** use SRM for all data transfer (default: False)
  `$ export DQ2_USE_SRM=True`
- **LCG_CATALOG_TYPE,** LCG catalog type
  `$ export LCG_CATALOG_TYPE=lfc`
- **DQ2_LFC_HOME,** LCG_HOME of local replica catalog
  `$ export DQ2_LFC_HOME=/grid/atlas`
- **DQ2_COPY_COMMAND,** which command is called in dq2_get. Specify this when srmcp doesn't work in your environment
  `$ export DQ2_COPY_COMMAND='lcg-cp -v --vo atlas'`

Available tools

- dq2_get
  - If a dataset is already present on a local storage element (SE) data is copied to the local directory or to another directory in the SE.
- dq2_ls
  - List information about datasets matching a given pattern. The pattern may contain wildcards which represent any strings. The wildcard symbol is the asterisk *.
- dq2_poolFCjobO
  - dq2_poolFCjobO resolves GUIDs for constituent files of a DQ2 dataset, and creates PoolFileCatalog.xml and Athena job-option.

Available tools (cont)

- dq2_put
  - Registers files to LRC, creates a dataset which is composed of the files, and then registers the dataset to DQ2. If a PoolFileCatalog is given, a list of files is extracted from the PoolFileCatalog. Otherwise, GUIDs are generated for files under a directory using uuidgen.
- dq2_register
  - It will upload local files to the Grid storage at the specified site and register them in the local Grid catalog and in DQ2 central catalogs.

Monitoring

- Panda dashboard:
  - http://gridui02.usatlas.bnl.gov:25880/server/pandamon/query?dash=ddm
- ARDA dashboard:
  - http://dashb-atlas-data-test.cern.ch/dashboard/request.py/site

Reference

- User Tools Twiki
  - https://twiki.cern.ch/twiki/bin/view/Atlas/UsingDQ2
- CVS
  - http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/DataManagement/DQ2_0_2/endusers/?only_with_tag=HEAD
  - (but check also - still empty) http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/DataManagement/DQ2/dq2.enduser/?only_with_tag=HEAD
- Generic DQ2
  - https://twiki.cern.ch/twiki/bin/view/Atlas//DistributedDataManagement
- Help (bug submission, troubleshooting)
  - https://savannah.cern.ch/projects/dq2-ddm-ops/
  - https://savannah.cern.ch/projects/panda/