



Recovery of Lost Files

Jiří Chudoba

26.1.2007

Institute of Physics, Prague



Scope

- Lost physical files on disks/tapes due to hw crashes, human errors, ...
- System Administrator provides a list of lost files
 - Production files and some users' files
 - names follow a convention (more than 1)
 - Files not following a naming convention
 - not covered here



Steps

- remove from the SE DB (system admin)
- delete lost entries from an LFC catalogue
- locate a replica
 - exists:
 - replicate
 - does not exist:
 - correct dataset (delete lost files)
 - pass list to prodsys



Update LFC: 1. Find LFN

/pnfs/grid.sara.nl/data/atlas/misal1_csc11/misal1_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misal1_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341._00380.pool.root.2

lfn:/grid/atlas/dq2/misal1_csc11/HITS/misal1_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misal1_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341._00380.pool.root.2

/pnfs/grid.sara.nl/data/atlas/calib0_csc11/calib0_csc11.005011.J2_pythia_jetjet.simul.HITS.v12003104_tid004283/calib0_csc11.005011.J2_pythia_jetjet.simul.HITS.v12003104_tid004283._00474.pool.root.9

/grid/atlas/dq2/calib0_csc11/calib0_csc11.005011.J2_pythia_jetjet.simul.HITS.v12003104_tid004283/calib0_csc11.005011.J2_pythia_jetjet.simul.HITS.v12003104_tid004283._00474.pool.root.9



Update LFC

- `lcg-lg --vo atlas`
`lfn:/grid/atlas/dq2/misall_csc11/HITS/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341._00380.pool.root.2`
`guid:52D6E5C4-B788-DB11-AD6C-0030485A052E`
- `lcg-uf 52D6E5C4-B788-DB11-AD6C-0030485A052E`
`srm://srm.grid.sara.nl/pnfs/grid.sara.nl/data/atlas/misall_csc11/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341._00380.pool.root.2`

Find replicas, replicate

- `lcg-lr --vo atlas`
`lfn:/grid/atlas/dq2/misall_csc11/HITS/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341.00380.pool.root.2`
`srm://se2.itep.ru/dpm/itep.ru/home/atlas/dq2/misall_csc11/HITS/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341.00380.pool.root.2`
- `lcg-rep -t 3600 -d`
`srm://srm.grid.sara.nl/pnfs/grid.sara.nl/data/atlas/misall_csc11/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341.00380.pool.root.2`
`srm://se2.itep.ru/dpm/itep.ru/home/atlas/dq2/misall_csc11/HITS/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341/misall_csc11.005106.PythiaWtaunu.simul.HITS.v12003106_tid004341.00380.pool.root.2`

Simple shell script to loop over all files, all 39 files replicated

It was easy because all replicas were in the same cloud



NIKHEF case

- 3604 files lost
- list of file names provided

```
144080 151  
/dpm/nikhef.nl/home/atlas/dq2/calib0/calib0.005011.J2_pythia_jetjet.simul.HITS.v12000301_tid003287/calib0.0050  
11.J2_pythia_jetjet.simul.HITS.v12000301_tid003287._00417.pool.root.14
```


- to get LFN:

```
sed 's#/dpm/nikhef\.nl/home#/grid#' lost_files.list > lost_files.lfc.list
```

- For some files need to add dq2:

```
sed 's#/grid/atlas/\.([^\dq2] [^\/*\])/#/grid/atlas/dq2/\.1/#' lost_files.lfc.list >  
lost_files.lfc.with_dq2.list
```

- lfc-ls \$FN
 - 18 minutes over 3604 files (3.3 Hz), all files found in LFC
- lcg-lg lfn:\$LFN
 - 21 minutes
- files were unregistered (loop in a shell)



Jiahang's approach

- decode **DS** name and **File** name:

[/dpm/nikhef.nl/home/atlas/dq2/calib0/calib0.005011.J2_pythia_jetjet.simul.HITS.v12000301_tid003287/calib0.005011.J2_pythia_jetjet.simul.HITS.v12000301_tid003287._00417.pool.root.14](#)

- find all sites with DS replicas
 - dq2.listFilesInDataset
 - find guid's for lost files
 - 3 minutes for 3604 files
-
- python script ***checknum.py*** in CVS
 - fast, but many files were not found in DQ2



Jiahang's approach (con't)

- Ifccheck.py
 - loop over sites (LFC's) where DS is registered and find registered replicas
 - Ifc_getreplica(",guid,se_host)

- Remove affected site from DQ2 DS registration
- Re-subscribe
- Clean DQ2 db if no replica exists



Work in progress ...

- short shell scripts capable to remove entries from LFC and replicate lost files
 - slow, but OK for short list of files
 - not general
 - not capable to replicate from other Grids
- more general and faster scripts in python are being developed