Files on disk

Files on disk on the Fermi Xrootd cluster

Infrequently the xrootd servers are scanned to find all the files that are on disk. The scan is split into directories below **/glast**:

> */glast/ASP, /glast/Data, /glast/bt, /glast/level0, /glast/mc, ...*

The summary shows the disk usage by subdirectories and duplicate files that were found.

The results of the scans are stored in

```
/nfs/farm/g/glast/u18/xrootd/files_on_disk/<YYYYMMDD>
```

where <YYYYMMDD> is the date the scan was taken.
There is also a link

> */nfs/farm/g/glast/u18/xrootd/files_on_disk/current*

that points to the most recent file listing.

File listings

For each subdirectory that is scanned the results are stored in the file:

filelist_<topdir>
    e.g.: **filelist_glast_Data** for files below /glast/Data

<topdir> is the directory the scan started and '/' are replaced by '_'.
Each line in the file listing contains the

```
size(Bytes)  atime  mtime  ctime  filename  host  [migrate-status]
```

The migrate-status field is optional with vaules:

status	description
-1	migrate status could not be determined (no lock file)
0	not migrated
1	migrated

e.g. :

> *107536534 1243879126 1217984515 1217984515 /glast/ASP/DRP/PROD/DRP_daily_00005.tar.gz*

The lines are sorted with respect to the data file names.

Summary Information

The file listing directory contains also summaries files that show the disk usage by top level directories and some sub-directories. The two summary files are:

- *usage_summary*
  usage for top level dirs (/glast/Data, /glast/mc,...) and some subdirs
- *usage_subdirs*
  usage for subdirs of /glast/Data/Flight, /glast/mc and /glast/mc/ServiceChallenge/

Duplicate files

There are files that exists on multiple file servers. A harmless case is where a file was copied from one to another server duplicating it. There are also files for which a file exists on two (or more) servers but their contents is different.

From the file listing the file names that exists on more than one server are extracted. For each of these file names we check if the size of the different instances are equal or differ.

The file **summary_duplicates** shows for the different FGST sub-directories how many of these files exists.
The **duplicates** sub-directory in the file-listing output shows the detail for all these files (dup_<path> files). It also contains the files named **notequal_dup_ <path>** that only shows those file names for which the file sizes differ.

Command used to collect the data

The command *SrvFlUsage* will create the file listing for each server and produce the summaries:
> SrvFlUsage -a -r
The listing directory will be /nfs/farm/g/glast/u18/xrootd/files_on_disk/<YYYYMMDD>