

## Intro

The goal of the project is to provide CDF community with means to stage data from CDF affiliated storage systems (FNAL dCache , CDF cnaf) directly to user's desktops. The goal of the project also requires that the provided solution must be available irrespective of the user's network location.

## Deliverables

### ***Sam get dataset***

The "sam get dataset" command is suggested as a top level user interface to stage data from SAM. The tool encapsulates diverse storage access protocols and guides users according to policies defined by CDF data handling.

Example usage:

```
sam get dataset --filelist=e8022932.0036exo5 --group=cdf --station=sam-public
```

### **Dependencies**

Sam station

Samcp

Jim\_gridftp

Vdt client

Gridftp server.

sam

### ***Station deployment***

Sam get dataset relays on SAM station to establish delivery of files. In addition SAM station is responsible for generating valid samcp source location for files that have been delivered to a storage. Since the convention to generate such location depends on the actual data access mechanisms, the specific SAM station instructions need to be set forth. These instructions enable SAM station to mangle the logical storage location path into samcp source location string.

- CDF central dCache server.

The central dCache logical storage path is the file's full pnfs location. I.E  
enstore:/pnfs/cdfen/filesets/./.

The cdfdca-ftp\* server pool will be re-used to access CDF dCache via gridftp protocol. Below is the list of configuration requirements to achieve that:

- SAM station disks :  
dcap://cdfdcache-ftp-backbone:cdfdca-ftp3.fnal.gov:2811/  
dcap://cdfdcache-ftp1:cdfdca-ftp1.fnal.gov:2811/  
dcap://cdfdcache-ftp2:cdfdca-ftp2.fnal.gov:2811/  
dcap://cdfdcache-ftp3:cdfdca-ftp3.fnal.gov:2811/
- SAM station consumption map and routing rules:  
--pmaster-arg=--consumption-map=\\.\\\*::dcap://cdfdcache-ftp1,  
dcap://cdfdcache-ftp2, dcap://cdfdcache-ftp3 --constrain-  
delivery=dcap://cdfdcache-ftp-backbone

- Gridftp server mount point (cdfdca-ftp1,2,3 is assumed as “/pnfs”. This assumption should be valid for CDF user principals only.
- SAM managed disk base storage.  
The SAM station disk logical storage location is <storage hostname>:<sam cache path>. Jim\_grid ftp server will need to be run manually on all nodes that host sam cache. Below is the is the list of configuration requirements:
  - SAM station disks  
No changes needed to an existing disk layout
  - SAM station consumption map and routing rules:  
Add --pmaster-arg=--consumption-map=sam\_download::<SAM node>::<SAM node>[:<jim\_gridftp server port nummver if available >]  
For every SAM node name in the cache.
  - Jim Gridftp server must be running on every SAM node name in the cache.

**Station version** : no restrictions

**Hardware:**

CDF production node : cdfsam02.fnal.gov

***Gridftp server deployment***

Grid ftp servers are essential part of the deployment. Grid ftp servers provide access to the data whether residing in SAM or the dCache managed storage elements in a secure and uniform fashion. Grid ftp servers are expected to be part of the experiment security infrastructure. We neither define nor enforce specific procedure that would be relevant to that.

**Gridftp version** : no restrictions

**Hardware:**

CDF gridftp servers : cdfdca-ftp1, cdfdca-ftp2, cdfdca-ftp3

***Client deployment***

[http://home.fnal.gov/~bartsch/sam\\_download](http://home.fnal.gov/~bartsch/sam_download)

sam version : v7\_5\_2

samcp version : current

jim\_gridftp : v2\_3\_5

vdt : v1\_1\_14\_13