gStore: Recursive File Handling

to facilitate handling of big file sets

gStore implementation (~Posix):

1. **recursive flags –r or –R**

   => same action for all (matching) files in top directory and all subdirectories

2. **wildcards in dirname do not replace ‚/‘, e.g.**
   - **/a*b matches /ab, /a1b, /axxb,**
   - **/a*b matches not /a/b**
3. **recursive handling and wildcards in dirname: several top directories, e.g.**
   - **/a*b/c matches /ab/c, /a1b/c, /axxb/c and subdirs**

gStore: Recursive File Handling

status:

- gstore **query**:
  - wildcarded dir names: from beginning
  - recursive file handling: since Feb 2011
- gstore **stage/unstage**:
  - since Feb 2011
- gstore **archive/retrieve**
  - in Jan 2012

gStore: Automatic Process Parallelization

past:

- **one copy process per cmd**
- **all data on same data mover**

now:

- **automatically several processes per cmd**
- **reading in parallel from different data movers**
- **writing in parallel to different data movers**

gStore: Automatic Process Parallelization

staging:

- **copy in parallel from n tapes to n data movers**
- **n<=8: limited by current no. of tape drives**
- **available since April 2011**

archive/retrieve:

- **n processes copying in parallel**
  - **between gStore disk cache and lustre**
  - **from gStore tape to lustre**
- **lustre: many OSSs**
- **planned availability: 2nd Qu. 2012**

gStore: Automatic Process Parallelization

advantage:

- **distribution of load over several DMs**
- **better utilization of available bandwidth**
- **shorter overall execution time**

valid for

- **copy process itself**
- **following data access**

gStore 2012/2013

offline Clients

lustre
>4 PB

Tape Roboter

5(10)GB/s

1.2 (2) GB/s

0.5 GB/s

capacity 2.2 PB
(8.8 PB)
expandable: 60 PB

online DAQ Clients

gStore Data Movers
200 (300) TB Buffer Storage