

MINERvA Software Infrastructure - Review Request #6140

Review Request # 6135 (New): Data Management Workflow Umbrella Task

Large Dataset Management

05/02/2014 10:32 AM - Gabriel Perdue

Status:	Assigned	Start date:	05/02/2014
Priority:	Normal	Due date:	09/02/2014
Assignee:	Daniel Ruterbories	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour

Description

Adopt the following procedures to manage large datasets in order to ensure efficient resource utilization:

- a. Work directly with the data storage group to enumerate the types and sizes of files that the experiment will store with the Enstore system. Establish file families for classes of files with dramatically different sizes or restore requirements. Continue to adjust the SFA policies, based on new optimizations from the data storage group, for each file family based on these sizes and recommendations from the storage group. Bypass the SFA system for files above the new thresholds established by the storage group to avoid future performance problems with restores from tape
- b. Perform staging requests through the SAM data catalog system or through special arrangements with the data storage group to ensure optimized restore of files from tape and to minimize the number of mounts per physical tape
- c. Reduce contention for disk/tape mover resources through administrative scheduling of tasks. Notify and schedule with designated CS-liaisons and the data storage group any large scale restore or reprocessing efforts prior to their start.

History

#1 - 02/23/2015 02:20 PM - Daniel Ruterbories

- Status changed from New to Assigned

- Assignee set to Daniel Ruterbories

Assign this to myself. Need to look at what families we have. If they are appropriate. Art Kreymer showed me how to check our families which I'm in the process of trying to do.