Strategies for Disk Usage

In Progress

Our current disk usage patterns are not viable going into the future. We have not budgeted enough money to buy disk at the rate we are using it. We (at least I) have blithely postulated in the past that we would find a magic bullet cut to "permanently" remove a large fraction of the downlinked data - which is ~>95% background. So far we have not done so, and there are no good ideas on how to yet.

Objective

I think our goal should be to have space for at least 2-3 copies of the data at any given time, and sufficient for a reasonable amount of MC.

Numerology

Early returns from survey running are that we consume 5 TB/week in L1. A snapshot of the L1 data so far shows:

Output File	% Total
Cal	14
Digi	9
FastMon	4
merit	3
Recon	60
SVAC	10

ignoring a bunch of miscellaneous files that probably eat up a few percent of the total. Recon is the biggest elephant in this room, but not by an order of magnitude. The main uses of recon are: TKR alignment and ToT calibration; event displays; and reprocessing.

From our most recent purchases, disk runs at 0.9k$/TB and tape at 0.4k$/TB.

Options

more disk and tape

2-3x data translates to $300-600k increment to current budget ($300k/yr disk+tape)

more tape only

keep cal, recon, svac and fastmon tuples on tape only (after some shortish period)

cost is $100-200k per year - ie $100k per copy of the data (or full reprocessing)

(Aug 18, 2008) We may want to think carefully before moving fastmon to tape, in case these files are useful for long term trending.

Delete recon

Sacrifice ability to reprocess quickly from recon. Changes reprocessing time from small number of weeks (1-2) up to a small number of months

10% solution

"permanently" removed 90% of non-digi/merit files from L1. The assumption is that tossed events will never (or almost never) be reclaimed from the garbage pile. Drops usage to 12% of current. All events are kept in digi and merit in this model.