

# WLCG Archival Storage Group

**Site Survey and “Making the most of tape”**

Oliver Keeble on behalf of the WG

# Archival site survey

- An attempt to establish best practise for optimally exploiting the archival storage systems used in WLCG
- A set of questions for each site designed to understand
  - Their advice to clients for efficient use of the system
  - The metrics available to spot anomalies and to track whether proposed improvements are effective
- Where possible, common tools such as FTS will be updated to follow the group's advice.
- <https://twiki.cern.ch/twiki/bin/view/HEPTape/Survey>
- Results - [https://twiki.cern.ch/twiki/bin/view/HEPTape/Survey\\_Results](https://twiki.cern.ch/twiki/bin/view/HEPTape/Survey_Results)

## Queue

What limits should clients respect?

---> Max number of outstanding requests **in number of files or data volume**

---> Max ~~requests submitted at one time~~ **submission rate for recalls or queries**

---> Min/Max bulk request size (**srmBringOnline or equivalent**) **in files or data volume**

Should clients back off under certain circumstances?

---> How is this signalled to client?

---> For which operations?

Is it advantageous to group requests by a particular criterion (e.g. tape family, date)?

---> What criterion?

## Prioritisation

Can you handle priority requests?

---> How is this requested?

## Protocol support

Are there any unsupported or partially supported operations (e.g. pinning) ?

## Timeouts

What timeouts do you recommend?

Do you have hardcoded or default timeouts?

## Operations and metrics

Can you provide total sum of data stored by VO in the archive to 100TB accuracy?

Can you provide space occupied on tapes by VO (includes deleted data, but not yet reclaimed space) to 100TB accuracy?

How do you allocate free tape space to VOs?

What is the frequency with which you run repack operations to reclaim space on tapes after data deletion?

## Recommendations for clients

Recommendation 1

---> Information required by users to follow advice

# Draft : Recommendations to experiments

- Submit recalls as far in advance as possible
  - Keep the queue as full as possible

## Campaign Planning

- Group recall requests by creation time or tape family if possible
- Inform the site with as much warning as possible about recall plans
  - Allows synchronisation with local activities such as repack
- Understand how priority requests are handled
  - Submitting priority requests will degrade throughput
  - Withholding recall submissions to keep latency down will degrade throughput
- Synchronise data use with recalls to avoid purge/recall loops
- The client should delete a staged file from the disk buffer once the workflow requiring the retrieval has completed.

## Client Behaviour

- Consider queue size to be unlimited
  - FNAL, PIC - 15k per VO
  - KIT - 2k per pool
  - UNIKHEF-SARA - 1k (?)
- Back off on SRM\_INTERNAL\_ERROR and SRM\_FILE\_BUSY
- Use bulk requests
  - Best bulk recall size unknown. 1k is the reference, some sites want more, some want fewer.
- Interaction rates under 10Hz typically acceptable
- Run with no timeouts, or at least 48hrs
- Ignore disk buffer occupancy
  - Exception: CNAF

# Discussion points on recommendations

- What payoff can we expect from the preceding advice?
- Should we make recommendations on writing strategy?
  - Selecting particular pools or resources for particular types of data? Probability of future delete??
  - Perhaps a writing strategy is not possible as repack will eventually destroy locality (?)
- More generally, the survey exposes the significant diversity in these systems. What should the strategy be?
  - Produce some "lower common denominator" advice - do the following... it may help, and will never hurt
  - Enumerate a small number of basic site characteristics and classify each site individually

# Next steps

- **Experiments have to join the conversation**
  - Which new scenarios to exploit tape are the experiments actually considering?
    - Carousels – R&D started at BNL
    - What else?
- Understand what actions, if any, need to be taken client side
  - FTS configurations can be updated easily
    - e.g. size of a bulk request, number of outstanding requests
  - What about developments in FTS or in experiment-specific clients?
- Track progress using the reported metrics
- Finalise the advice to clients in a “mode d’emploi” for tape systems
- Other activities
  - Investigate buffer dimensioning and drive allocation strategies
    - RAL & PIC at least
  - Make contact with the “technology watch” WG
- Next meeting to be scheduled shortly, with experiment involvement

# Summary

- The Archival Storage Working Group is trying to describe *how* to use tape systems effectively
- The experiments now have to describe *what* they want to do
- And we have to try to do it
  - Using the metrics to see what happens