LPC Data Organization

Robert M. Harris
Fermilab

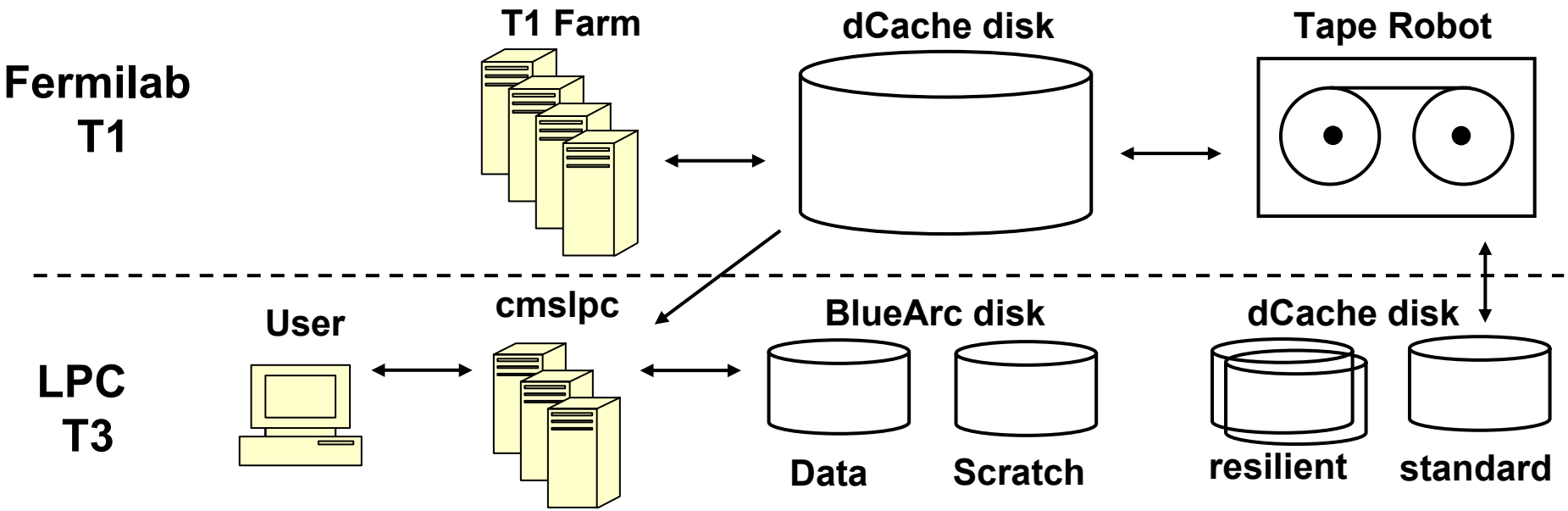
With contributions from
Jon Bakken, Lothar Bauerdick, Ian Fisk,
Dan Green, Dave Mason and Chris Tully.

USCMS Run Plan Workshop
May 16, 2008

Introduction

- LPC has dedicated data management resources
  - Distinct from the Fermilab T1
- Here we give a brief overview of those resources
  - The “type” and amount of disk
  - How it can be accessed and used
- Begin the discussion of how to allocate the resources
  - Present the initial plans of LPC and computing management
  - Solicit the input of the LPC users and USCMS
- These are your resources and we want your input
  - What are your data management needs and issues at the LPC ?

T1 and LPC Disk Use



- T1 data storage resources are transparently available to LPC users
  - Cmslpc can access data on T1 disk via dCache. Available for analysis.

- Dedicated LPC disk resources come in multiple types
  - BlueArc disk: not backed-up but easy to use and to allocate to users and physics groups.
  - dCache disk:
    - Resilient dCache: protected by 2nd copy on disk (resilient). Able to hold small user files.
    - Standard dCache: proetcted on tape, but only for data samples with large files (semi-official data).

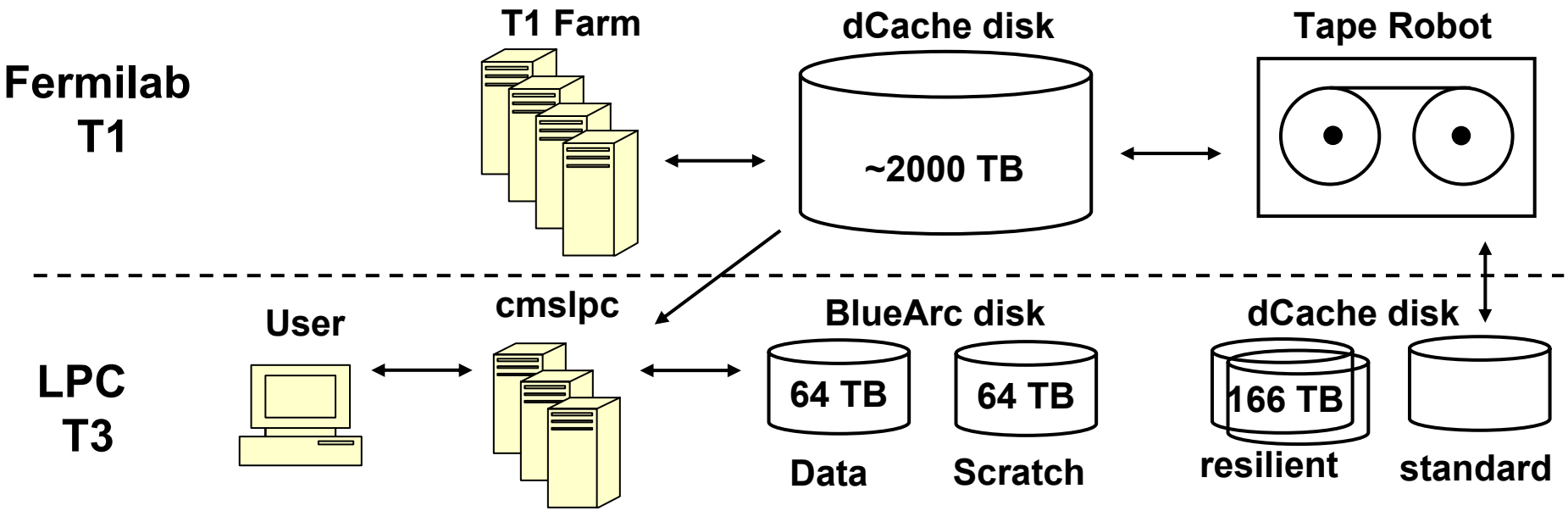
dCache data pinned on disk

- dCache is normally a disk cached system of tape access.
  - Data goes to and from tape through a disk cache.
    - Data does not “stay” on disk normally.
  - Recently accessed data is often on disk and accessible quickly
    - Older data will often take longer to access off of tape.

- dCache data can be pinned on disk.
  - The data will then stay on disk for the lifetime of the pin.
  - dCache access then does not have any delays from tape access.
    - Data is still protected by being on tape.

- We can pin critical datasets on disk for access w/o tape delay.

T1 and LPC Disk Currently



- T1 disk (2 PB) can hold all AOD and FEVT 2008 run data pinned on disk.
  - LPC users can then transparently access all data (AOD, FEVT) on disk quickly.
- LPC disk for user data and MC is now 128 TB not counting resilient dCache
  - Data: /uscms_data/d1 is 64 TB, 11 used, and 51 allocated. 20 GB/user default
  - Scratch: /uscmst1b_scratch/lpc is 64 TB, 51 used, no quotas.
  - Resilient dCache: 166 TB, 86 used (holds 43), 80 available. No quotas now.

LPC disk use proposal

- The LPC disk will grow significantly to reach TDR goal of 1 TB / year / user
  - Current plans are to add 384 TB.
- BlueArc
  - The data area would be tripled from 64 to 192 TB.
    - 64 additional TB for physics groups.
    - 64 additional TB for users.
  - Up to 1 TB / user on request
    - With the default for users increasing from 20 GB to 100 GB
  - The scratch disk size remains 64 TB
- dCache
  - Tape-backed dCache with up to 256 TB pinned on disk for rapid access
    - Physics group and user skims, compressed datasets, and MC datasets.
  - Resilient dCache remains 166 TB with quotas of 1 TB/user soon.
- In this proposal all users get a lot more disk
  - All users get 100 GB in BlueArc, 1 TB on request. 1 TB/user resilient dCache.
  - Datasets larger than this would be in dCache backed to tape.

Conclusions

- We should have plenty of disk for the 2008 run
  - Run AOD and FEVT data should be pinned on disk on T1
    - Transparently accessible from the LPC
  - Other priority datasets can be pinned on disk at the LPC
    - Priority skims, compressed datasets, MC datasets.
- We will have a lot of disk for LPC users
  - More than half a PByte for user files and datasets.
- We've presented some initial plans
  - We welcome your comments.

Backup Slides

Data Management Issues

- Data Transfers
  - Users may request transfers of datasets to the LPC
    - Currently they go to the T1, but in the future they will go to the LPC.
    - Transfers now go into dCache backed by tape.
    - Transfers must be approved by data manager. Rejection criteria?

- Users writing to dCache
  - Users can only write to resilient dCache
    - Avoids writing of small files to tape robot which hurts tape drives.
    - Currently mostly done via CRAB.
    - Grid credentials required for writing of files via SRMCP.
      - **Higher "barrier to entry" may be a good thing.**
  - "Official" datasets can be migrated to tape-backed dCache.