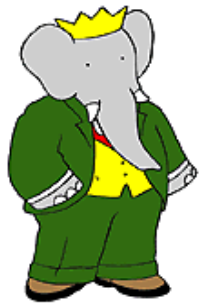




CCRLRC
Rutherford Appleton Laboratory



TM and © Laurent de Brunhoff

Data Distribution and Management

Tim Adye

Rutherford Appleton Laboratory

CM2 Implementation Oversight

26th June 2003

Overview

- July Test Production
- PR and SP
- Imports to Tier A/C
- XRootd

People:-

Tim Adye (Kanga, Grid), Dominique Boutigny (Grid),
Chris Brew (Kanga), Alasdair Earl (Grid),
Alessandra Forti (Bookkeeping),
Andy Hanushevsky (XRootd, Grid), Adil Hasan (Grid),
Wilko Kroeger (Grid), Liliana Martin (Grid),
Jean-Yves Nief (Grid, XRootd), Fabrizio Salvatore (Kanga)

July Test Production

- Will use existing skimData database
- Just need to handle multiple files per collection
 1. Add a new `skim_streamstats` table column with a list of file extensions
 - Alessandra Forti will update table schema and fill data from PR (with help from Teela Pulliam) and skim production
 2. Update `skimImport` to import all files from each collection
 - Actually allow global selection of mini vs micro at command-line
 - Tim Adye will make this change as soon as (1) is complete

Limitations of July version

- Don't plan to include user data
 - could add this if needed
- Import sites can only select mini vs micro globally when files are imported
 - Only one `import_status` field per collection
 - Sites can work around this restriction, but it's messy

PR and SP exports

- Sorted out much of the architecture at Lyon
 - PR/SP exports will still be a “push”
 - Imports to other Tier A/C is “pull”
 - Merging and integrity (RFI) check done at production site (at least logically)
 - Three-step publishing process
 - Registration - add db entry
 - Location - move to SLAC or other QA site
 - Publication - ready for physics analysis
- PR and SP export process **much** simpler than for Objectivity
 - SP: no longer need MocaEspresso: Douglas Smith aims to “borrow” code from skimImport
 - PR: stripped down version of existing Padova export

Exports to Tier A/C

- Future changes will mostly concentrate on the new bookkeeping
 - Use file catalogue: gives collection to logical-filename mapping
 - Table definitions and API
 - Dataset bookkeeping tables (distributed everywhere) will just say whether a collection is at each Tier A
 - Each file needs its own local-status flag
 - This will simplify existing distribution and management tools
- Finalise new table definitions once July production is under way
 - Tim Adye and Alessandra Forti

XRootd

- We compiled a long list of XRootd requirements
 - Artem will discuss these with Andy
- XRootd will only be used for reading
 - NFS/bbftp/... for writing
 - At least for the moment
- Schedule:-
 - XRootd-enabled release (**XNetFile**) will be available this week
 - Jean-Yves (ccIN2P3) and Chris Brew (RAL) will start testing: Chris has started, Jean-Yves will in July
 - XRootd with load balancer available by end July

Conclusions

- Only minor change required for July test production
- PR and SP exports much simplified by removal of Objectivity
- Tier A/C imports simplified by new bookkeeping tables
- Testing XRootd now. "Full" version available end of July.