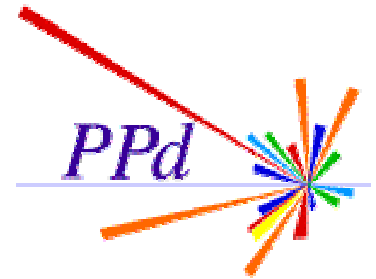


TM and © Laurent de Brunhoff



# Data Distribution

Tim Adye

Rutherford Appleton Laboratory

Augmented Technical Board Meeting

6<sup>th</sup> April 2001

- Status of **Kanga** export tools
- Status of **Objectivity** export tools
- Objectivity data distribution **developments**
- **Grid**
- Summary

# Kanga Status

- New Kanga distribution system **complete** and **in use** at various sites [A.Forti, TA]

- RAL, (Rome), Manchester, Bristol, IC, Colorado State,...
- Uses local copy of **skimData** database
- **Automatic** operation
  - cron or repeating batch job
- **Documentation** just released

- Kanga **backup/archive** procedure being commissioned [A.Dorigo]

- Under test in Rome
- Requires **local customisation** for different tape systems
  - eg. HPSS at SLAC
- No automatic **staging** system yet

# Objectivity Status

- Outside SLAC, Objectivity used for
  - Full micro analysis at Lyon
    - Will expand as Tier A Centre ramps up
  - Dominique's talk
  - SP production
  - Fergus' talk
  - Some private MC production
    - Not exported
  - Some small-scale tests and preparation for the future
    - Eg. Andy Hanushevsky set up latest secure versions of AMS / oofs / ooss and PUD at RAL
- To implement **Computing Model**, need to support **widespread** distribution and use of Objectivity

# Current Developments

Fill **collection metadata** for all new data

[M.Marzolla]

- Tells which database files store a collection's data
  - Otherwise require a lengthy scan
- First **tests** on SP4 production data look **OK**
- Needs to go into **OPR** and **SP** production
  - Or downstream, but probably before writing to HPSS
- Cannot do this for existing data
  - Full scan would require staging in all **raw+rec**
  - Maybe partial scan would still be useful
  - Will be done automatically for staged data (**collstagein**)

# Current Developments

## Use new bulk data transfer tools

- [bbftp](#)
- [bbcp?](#)
- [GridFTP?](#)

[TA]

[G.Farrache, IN2P3]

[A.Hanushevsky, SLAC]

[GLOBUS]

# Remote Simulation Production

- MocaEspresso **automates** SP export and transfer to SLAC [D.Andreotti, E.Leonardi]
  - Exports “full” databases while production continues
    - **Faster** turnaround
    - No more need for large staging disk at each site
  - No longer need to be an expert on BdbDistTools arcana
- **Automatic** SP import procedure at SLAC [C.Bulfon, L.Mount, A.Hasan]
  - Minimal manual intervention at SLAC end
  - In production

See Fergus' talk

# Bulk Data export

- BdbServer exports data from SLAC [D.Boutigny, A.Zghiche]
  - Automatic or on request
  - See Dominique's talk
- Investigating whether skimData can be used for finer control of what is exported [B.Jacobsen]
  - Analogous to new Kanga export scheme

# Multiple Federations

- Multiple federations should greatly **simplify** data distribution
  - Removes **DBID limit**
    - Can write **smaller** DB files
    - Can add more **streams**
    - Reduce unwanted data in export
  - Improves modularity of import and export
- ... in the long term. **Now** we need to
  - add support to data distribution tools
    - Awaiting Bdb support
  - treat each federation separately for the moment

# Gridery

- Lots of development work (in BaBar and outside)
  - Storage Resource Broker (SRB) from SDSC [R.Schmitz]
  - Globus data management tools (GASS)
    - Both provide a catalogue of distributed data + ...
  - GDMP
    - Objectivity data replication from CMS
  - LDAP interface to skimData?
  - ...

# Summary

- **Kanga** data export is up and running
  - Backup and archiving tools under test
- **Objectivity** developments under way
  - Add **metadata** to public federations
  - Improve data **transfer rates**
  - Automate **SP** export / import
  - Automate **bulk data** export / import
  - Add support for **multiple federations**
- Keep abreast of **Grid** efforts