Data Distribution

Tim Adye
Rutherford Appleton Laboratory

BaBar Collaboration Meeting
15th December 2000

- SP Exports
- SP Imports
- Objectivity Data Exports
- Kanga Data Exports

Lots of other excellent and vital work going on behind the scenes…

SP Exports

- SP data generated at remote sites has to be exported from site Production Federation, copied to SLAC, and then imported to SLAC Federation

- Old system manual, tedious, and error-prone
  - Production stopped
  - Arcane BdbDistTools commands executed
  - Long wait for these to complete before restarting production
  - Files copied to SLAC

Automatic SP Export

- New tool by Emanuele Leonardi & Daniele Andreotti
- Prototype tested at a few sites
  - Automated export (standard BdbDistTools) and ftp
  - GUI control
- New version performs export in parallel with production
  - Only “closed” (full) databases exported
  - Needs less staging space
    - 100+ GB -> 20 GB
- Transfer to SLAC now uses bbftp.
- Will be tested on next Production cycle at Rome

SP Imports

- Production sites copy their data to datamove3 for import into SLAC Federation

- Import procedure now automated [Cristina Bulfon]
  - Runs from a cron job
  - Checks for new export from Production sites
  - Maintains e-logbook of operations [Lawrence Mount]
- Requires production sites to follow simple protocol
  - See DataDist HN 111 for details

- Still need to improve error handling

Objectivity Data Exports

- Automated bulk exports to IN2P3 continue
  - Further efficiency improvements introduced

- BdbServer, bulk export tool can also simplify small user exports:
  - Bypasses many common problems
    - Finding suitable export space (and tidying it up afterwards!), authorisation problems, complex options
  - Simple e-mail interface
    - Send list of DB IDs
      - (eg. from colldb list; metadata list in future)
    - Data placed on ftp-accessible disk
  - See Data Distribution Web page for details

Kanga exports

- 228k Kanga files, using 4 TB – and still growing!
- Current procedure (syncslac / rsync) is too slow ☹
  - >4 hours, just scanning directories for new files
  - Data transfer not optimised for WAN
- New procedure [Alessandra Forti, TJA]
  - Uses skimData catalogue to find new files (~10s)
  - Uses optimised “ftp” tools (bbftp / sfcp)
    - Larger TCP/IP window size
    - Multiple streams for each file
  - Eg. bbftp SLAC<->RAL gives x10 improvement! ☺
    - Your mileage may vary – a lot

Kanga export tools - Status

- Tested SLAC->RAL: 2 TB -> 3.1 TB
- A couple more sites recruited as “guinea pigs”
  - More welcome!

- New tools for local management under development

[Alvise Dorigo]

  - Backup/archive to tape
  - Delete old files
  - Controlled by changes in skimData database

Conclusion

- Automation and efficiency improvements
  - SP Exports – under test
  - SP Imports – in production
- Simplified small user Objectivity exports
  - Available for use
- Much faster Kanga export procedure
  - Being deployed