PROOF Status and Roadmap

Fons Rademakers, Gerri Ganis, Jan Iwaszkiewicz
CERN

ALICE Offline Week, April 9 2008

Current Issues

- Many instabilities due to:
  - Deadlock in xrootd and xrdproofd
    - Fixed by refactorization and revision of locking strategy in xrdproofd
  - Missing files not properly handled in workers
    - Fixed in workers, additional fix in packetizers following
  - Failing writev() in xrootd causing connection to be closed
    - Causing 5 min wait in data server
    - Causing dying of workers
    - Fixed
  - Startup delay due to unavailable hosts
    - Need dynamic host list
  - Crash due to double delete of merged histograms in analysis framework

New Features Under Test

- Priority based scheduling
- File quotas
- More extensive xrootd file access monitoring

Planned Developments

- Error handling when a worker dies
  - Contact lost, no ping
  - Resubmit packets on other workers
  - Need redundant files

Planned Developments

- Error handling when a worker segv's in user code
  - Send in error handler which packet and event was being processed
  - Resubmit all packets minus that one event
  - Report to client which exact event caused problem

Planned Developments

- Scheduling interface with Condor
  - Condor scheduler controls PROOF sessions
  - Joint project with US ATLAS/Condor team Wisconsin/PROOF team

Planned Developments

- Simplify installation and configuration
  - Zero configuration
  - Dynamic host information

Planned Developments

- Increase scalability by using sub-masters to parallelize merging
  - Depending on the number of workers create a sub-masters
  - Each sub-master controls a number of workers
  - Sub-masters do first level of merging of the worker results

Planned Developments

- PROOF Lite
  - PROOF optimized for multi-core systems
  - Client talks to workers, no master, unix sockets
  - Memory mapped file access

Planned Developments

- Increase monitoring and feedback capabilities by adding a lightweight httpd thread in the master
  - Allows getting progress histograms and process information (number of packets processed, number of failures, etc.)
  - Select and view (pull) feedback histograms only when needed instead of always pushing them to the client

Conclusions

- PROOF parallel performance good
- Stability problems have highest priority
- Open bugs are being addressed
- Increasing interest
  - US ATLAS scheduling wide deployment in Tier-3s
  - CMS several sites running tests
  - LHCb waiting for Python based TSelector support
    - First version of TPySelector implemented