

# PanDA and panda-client News

Tadashi Maeno (BNL)

# Outline

- Automatic freezing of output datasets
- Integration of Event Picking to PD2P
- --rootVer option of prun
- AMI tag option for Event Picking
- Automatic job splitting to avoid jumbo output datasets
- New output file merging scheme
  - See Hurng-Chun's talk

# Automatic freezing of output datasets

- DQ2 doesn't transfer user files until datasets are frozen
  - All user datasets are frozen by DQ2 after 7 days → Users had to manually freeze datasets if they wanted to transfer files more quickly
  - Previously dataset had to be open
    - New files were added to dataset when the dataset was reused
- Now the default output of analysis job is container instead of dataset
  - When container is reused, new dataset is added to container every time new job is submitted → old dataset can be frozen
- The panda server freezes dataset when all jobs contributing the dataset have finished/failed/canceled → DaTRI can transfer files more quickly

# Event Picking in PD2P

- To transfer only files which contain interesting events, rather than transferring the whole dataset
  - Proposed in the context of life without ESD
- New functionality added to PD2P
  - Takes a list of event and run numbers, converts them to a list of files using EventLookup service and DQ2, makes a temporary dataset
  - The dataset could be used for data transfer
- Integration with DaTRI or client tools?

# --rootVer option of prun

- Normally users use ROOT included in AtlasRelease
  - The pilot sets up Athena on WN, then ROOT is available
- Sometimes users want to use newer version of ROOT which is not included in AtlasRelease yet
  - Bugfix, new functions, etc
  - Release cycle of AtlasRelease is not synchronized to ROOT's one
- --rootVer option is added to prun since 0.3.28
  1. Pre-compiled ROOT is downloaded from the panda server to WN and is included in lib.tgz in the build step
  2. ROOT is set up on WN in the run step, so that user's application can run on the top of the ROOT
- **Caveat**
  - Pre-compiled ROOT requires gcc4.3. The easiest way to have gcc4.3 on WN is to setup Athena
    - --athenaTag option is required together with --rootVer option

# AMI tag option for Event Picking

- Event Picking relies on EventLookup service which is provided by TAG team
- EventLookup service uses the latest version of AMI tag by default
- Sometimes users are interested in old AMI tag
  - e.g., first pass reconstruction → f\*
- `--eventPickAmiTag` option is available since 0.3.30
  - Allows users to run on data with specific AMI tag

# Automatic job splitting to avoid jumbo output datasets

- A user produced very large dataset which contains ~200k files
  - Problematic for DDM and SE
- DQ2 has imposed a limit on the number of files in a dataset (20k)
- pathena/prun has been updated to split jobs automatically, so that dataset\_N (N=1,2,3...) is created per 10k output files or 4k subjobs and those datasets are added to output container
  - Transparent to users as long as output container is used (default)