

Pegasus 4.0

Table Of Contents:

- 1 Main Focus
 - 1.1 Testing Framework and Testing
 - 1.2 FHS Work
 - 1.3 Pegasus Lite and Improving Condor IO Staging Mode
 - 1.4 MonitorD
 - 1.5 Stampede Related Activities
 - 1.6 Other Pegasus Changes
 - 1.7 Other Things
 - 1.8 Porting the VM to 3.2.0

Last Updated September 30th , 2011

Main Focus

Testing Framework and Testing

- Jens and Gaurang will work on setting up a Bamboo Framework

FHS Work

- Main JIRA item
 - <https://jira.isi.edu/browse/PM-488>
- Moving pegasus to native Debian and Redhat packaging
- Change of all tools to pick up new layout.
- kickstart/seqexec/keg to be renamed to have a pegasus prefix
- Planner code needs to be changed to reflect new naming and layout.

Pegasus Lite and Improving Condor IO Staging Mode

- Main JIRA item
 - <https://jira.isi.edu/browse/PM-493>
- Notion of Staging Site in the Planner
- Staging of worker packages for non shared filesystem
- Pegasus-Lite- Shell launcher for Jobs on remote sites
- New clients for create dir and cleanup jobs that can run locally?
 - for create dir , transfer of zero byte file
- Do away with the S3 specific implementations
- Do away with Condor Refiners for Condor IO and fold it to the SLS interfaces.

MonitorD

- refactoring of monitorD code. right now it is all in one file!
- Locking mechanism to ensure that multiple instances of monitorD cannot run on same submit directory
- boolean option whether to store stdout or stderr in the database

Stampede Related Activities

- Main JIRA item
 - <https://jira.isi.edu/browse/PM-504>
- Porting of pegasus analyzer
- Reading of DAGMan generated jobstate.log files
- pegasus-statistics changes
 - Option to expose stats as a csv file?
 - pegasus-statistics should take in multiple workflow id's / pass the list of directories
 - statistics . size of task edge tables. can the size be decreased. we need to talk to stampede folks more about it.
- New Plots and Statistics?
- CPU Time tracking in addition to the walltime (as currently reported by kickstart).
 - - Jens says utime is in the kickstart record
 - should we change the schema ?
- Upgrading the stampede schema's
 - how do we handle the upgrade to handle extra things
- multicore jobs calculations.
 - there should be a multiplier /multiplying factor that is propagated through the system.

- pegasus can put the number of cores in the submit file
- storing status vs exitcode
 - if a job dies on a signal , can we have a special exitcode so that kickstart record has an exitcode
 - in the stampede schema , monitor should only populating raw status in the db
 - pegasus analyzer should be smarter about whether failure happened because of signal and or a job failed by exitcode.
 - put in an extra column in the stampede schema ???

Other Pegasus Changes

- Hierarchical Workflows
 - Propagation of options from command line to the sub workflows.
- Time based clustering
 - Similar to horizontal clustering.
 - runtime profile key that can be used with the jobs.
 - another profile key that says the clusters.maxruntime.

Other Things

- Option to do a mv on stageout . Brain workflows usecase.
 - to do in the transfer client not necessary in the planner. needs more thought and discussion

Porting the VM to 3.2.0

People Involved [Karan,Rajiv]

- Addition of new exercises
 - How to add notifications