

Pegasus 3.1

Table Of Contents:

- 1 Main Focus
 - 1.1 JIRA Task Board
 - 1.2 Stampede Related Changes
 - 1.3 Notification Support in Pegasus via monitord
 - 1.3.1 Monitor Management [Fabio,Gaurang]
 - 1.4 Auxillary Tools to Stampede DB
 - 1.5 Monitor Changes [Fabio]
 - 1.6 DAX API changes
 - 1.7 S3 support [Mats, Karan]
 - 1.8 Usability Changes
 - 1.8.1 Condor Common Log Handling?
 - 1.8.2 Open Question
 - 1.9 User Guide Reorganization [Bill]
 - 1.10 Testing Framework and Testing
 - 1.11 Porting the VM to 3.1.0
- 2 Timeline/ Sequence of changes
 - 2.1 For Stampede related changes
 - 2.2 Notification Related Changes

Last Updated April 5th , 2011

Main Focus

- Add notification support to Pegasus
- Moving auxillary tools to stampede db
- Other Stampede Related Changes
- Usability Changes
- User Guide Reorganization
- Testing Framework and Testing.

JIRA Task Board

[TaskBoard for 3.1](#)

[Items for April End](#)

Stampede Related Changes

- Main JIRA Issue
 - [Main JIRA Item PM-345](#)
- Updated Database Schema
 - [Schema Page](#)
- Additional DB schema changes to be able to connect jobs/tasks in the DAX with corresponding kickstart records [Karan,Fabio,Monte]
- Identify the new events for the updated schema [Karan,Fabio]
 - Pegasus to generate netlogger events to a file in the submit directory [Karan]
 - Changes to monitor to conform to new schema, populate netlogger stream from Pegasus. [Fabio]
- Addition of workflow metrics file containing distribution of jobs into the DB [Fabio,Karan]
 - Revive the metrics file created by Pegasus . Should be populated in the submit directory. [Karan]

Notification Support in Pegasus via monitord

- Main JIRA Issue
 - [Main JIRA Item PM-346](#)
- Monitor needs to be managed by Condor. What happens if monitor crashes or condor/system crashes? We want monitor to come up automatically as Condor recovers after a restart. [Fabio,Gaurang]
- Monitor needs to support notifications [Fabio]
- Requires changes to Pegasus to generate input file for monitor [Rajiv,Karan]
- Come up with default notify scripts in the toolkit that notify the user and generate some status reports. [Gaurang]
- Changes to DAX Schema
 - [Main JIRA Item PM-350](#)
 - Addition of invoke element at the workflow level
 - Changes to python API [Gideon]
 - Changes to Perl API [Jens]
 - Changes to JAVA API [Gaurang]
 - Change to JAVA Parser [Karan]

- Fabio needs to make sure exitcodes are thrown correctly and restarts are handled correctly. **[Fabio]**

Monitord Management **[Fabio,Gaurang]**

<https://confluence.pegasus.isi.edu/display/pegasus/Monitord+Management+via+Condor>

Auxillary Tools to Stampede DB

- pegasus-statistics **[Prasanth, Mats]**
- pegasus-plots **[Prasanth]**
- pegasus-analyzer **[Fabio]**

Monitord Changes **[Fabio]**

- Monitord also needs to be able to account for newer versions of Condor DAGMan creating a jobstate.log file.

DAX API changes

- Feedback from Duncan brown while using Python API
- Executable handling in Python API and JAVA API
 - internally executables handled as lists not as Sets
 - python api does not allow for adding edges based on id's.
 - also have a getJob function based on ID.
 - escape function in python api?is it for all strings
- JAVA API **[Gaurang]**
- Python API **[Gideon]**
- Perl API **[Jens]**

S3 support **[Mats, Karan]**

- pegasus-transfer to support pegasus-s3 tool Gideon wrote **[Mats]**
- changes to Pegasus to use pegasus-transfer for S3 **[Karan]**
 - handle transfer of S3 config file etc
 - get s3 to work with SeqExec launcher.

Usability Changes

- refactoring of clustered jobs (internal to pegasus)
- Addition of -conf option
 - Java Clients **[Prasanth,Rajiv,Karan]**
 - Python Clients **[Prasanth,Fabio]**
 - pegasus-statistics, pegasus-plots, pegasus-analyzer, monitord
 - Perl Clients **[Gaurang]**
- Improvements to pegasus-tc-client **[Prasanth]**
 - the pegasus-tc-client output is in the old deprecated format.
- Improvements to pegasus-rc-client **[Rajiv]**
 - Investigation of RLS compatibility issues
- Addition of default categories to allow for easier specification of category based knobs at DAGMan level
 - cleanup jobs
 - subdax jobs
- Improve Rescue DAG semantics **[Rajiv]**
- Pegasus should not require the jobmanager compute to be present in site catalog. **[Karan]** <https://jira.isi.edu/browse/PM-277>

Condor Common Log Handling?

- Condor Common Log Handling to be discussed. <https://jira.isi.edu/browse/PM-222>

Open Question

Improve the Condor File IO mode in Pegasus ? Not clear how to do it without going down the staging-sites option. User experience can be improved, but would be a hack to do it in Pegasus without staging-sites option.

User Guide Reorganization **[Bill]**

- Dependant on Bill

Testing Framework and Testing

People Involved [Jens,Gaurang]

Porting the VM to 3.1.0

People Involved [Karan,Rajiv]

- Addition of new exercises

Timeline/ Sequence of changes

For Stampede related changes

1. Stampede DB schema redesign and then support needs to be implemented by Fabio and Monte.
 - a. Original estimate by Fabio for db changes and porting of pegasus-analyzer was end of April.
 - b. Depending on scale of schema redesign, this **may have to be extended** !
2. Only after that pegasus-analyzer, pegasus-statistics and pegasus-plots can be ported for 3.1.

Notification Related Changes

1. First monitor needs to be managed
2. Create input file for monitor containing the notifications for the jobs.
3. Fabio puts in support for notifications in monitor.

Fabio should first work on monitor to be managed. Then move to stampede changes.

This tells us up front if monitor can do notifications. Assumption is that it is critical to not have missing notifications in case of system crash.

If we are ok with [as is approach](#) , then management of monitor is not an issue.

While Fabio works on stampede, we can do the creation of the input file for monitor.

If no major stampede db changes are decided, then following might be feasible

- Fabio gets stampede related changes done end of april.
- Fabio gets whole of may to put in notification support.
- End of May we may have a first beta, where everything identified above is done.
- June is spent testing on the release .