

# Pegasus 3.1 Original Plan

## Table Of Contents:

- 1 Main Focus
  - 1.1 Data to Worker Nodes
  - 1.2 Notification Support in Pegasus via monitord
    - 1.2.1 Open Question
  - 1.3 Auxillary Tools to Stampede DB
  - 1.4 Monitor Changes [Fabio]
  - 1.5 Stampede Related Changes
    - 1.5.1 Usability Changes
  - 1.6 User Guide Reorganization [Bill]
  - 1.7 Testing Framework and Testing
  - 1.8 Porting the VM to 3.1.0

Last Updated March 17th , 2011

## Main Focus

- Improving how we get data to worker nodes. Being able to retrieve data on the worker nodes from various sources.
- Add notification support to Pegasus
- Moving auxillary tools to stampede db
- Other Stampede Related Changes
- User Guide Reorganization
- Testing Framework and Testing.

## Data to Worker Nodes

The main change for 3.1. The developers document has complete details  
[Download Developers Document](#)

- Refactor how clustered jobs are handled [Karan]
- Addition of staging-sites option to pegasus-plan [Karan]
- New Shell Job Wrapper for jobs when running on worker nodes [Karan,Mats]
- Reconciling the Condor File Transfer Mode [Karan,Mats]
- Changes to pegasus-cleanup
  - Creation of pegasus-cleanup jobs [Karan]
  - new pegasus-cleanup client that runs locally [Mats]
- Changes to pegasus-dirmanager
  - New pegasus-dirmanager client [Mats]
  - New java COG client called pegasus-gridftp client that will create directories and remove files [Gaurang]
- Special S3 support into pegasus-transfer, pegasus-cleanup and pegasus create dir [Mats]
- Bypassing of staging-site while staging in input data [Karan]
- Bypassing of staging-site while staging out output data [Karan]
- Transfer of braindump file to remote workflow execution directories [Karan]
  
- Open Task. Not sure.
  - Pegasus will also add leaf remove directory jobs that remove the workflow execution directory from the staging site to the executable workflow [Karan]

## Notification Support in Pegasus via monitord

- 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
  - 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]
- Monitor needs to be managed. What happens if monitor crashes or condor/system crashes? We want monitor to come up automatically as Condor recovers after a restart.
  - Instead of pegasus-run launching monitor, monitor should appear as an independent job in the workflow with the highest priority [Karan]
  - Fabio needs to make sure exitcodes are thrown correctly and restarts are handled correctly. [Fabio]

## Open Question

- Notifications are required at the workflow level. But how does it affect the DAX/DAG jobs ? In the parent workflow, the dax and dag jobs are jobs, and at same time they have separate sub workflows associated with them. So notifications for a DAX/DAG jobs in the parent workflow will clash with workflow level notifications in the sub workflows.

## Auxillary Tools to Stampede DB

- pegasus-statistics **[Prasanth]**
- 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.

## Stampede Related Changes

- Improve Rescue DAG semantics **[Rajiv]**
- Additional DB schema changes to be able to connect jobs/tasks in the DAX with corresponding kickstart records **[Fabio,Monte]**
- 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]**
  - Monitord picks up the metrics file and stores it in the DB **[Fabio]**

## Usability Changes

- 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]**
- Improvements changes 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

## 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