lmsim On Demand

DESC Collaboration Meeting -- June 2014

Jaroslava Schovancova, Torre Wenaus (BNL)
Debbie Bard, Jim Chiang, Seth Digel, Richard Dubois, Warren Focke, *Tony Johnson*, Brian Van Klaveren (SLAC)

Contents

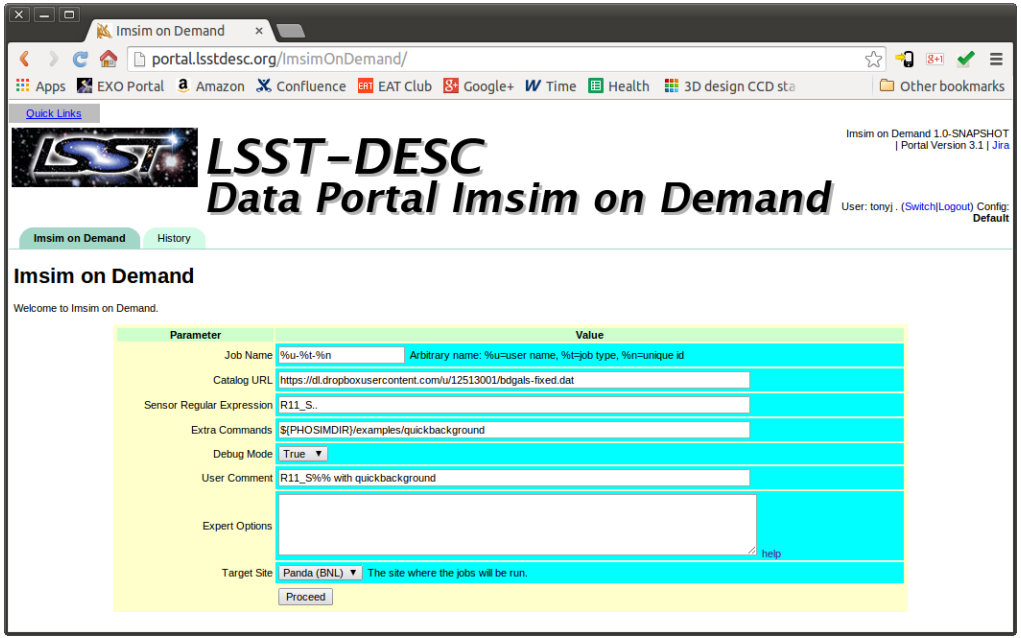
- Current goals of “lmsim on Demand”
- What currently exists (proof of principle)
- Technologies used
- Future plans?
- Request for input

Imsim On Demand goals
(Richard Dubois @ Dec '13 DESC meeting, Pitt)

- PhoSim is easy to set up and run on a small system
  - Much harder to manage production on thousands of cores and worse on big MPI machines
- Work towards a simple web interface for running ImSim on demand, supplying a catalogue and PhoSim parameters
  - Fire off jobs into the ether and notify the user when done (or if action is needed to rerun any failed jobs)
- Make use of existing workflow tools
  - Workflow engine to run graph of jobs
  - data catalogue to record where the data went
  - Interfaces to GRID, HPC, clouds etc to find the cycles

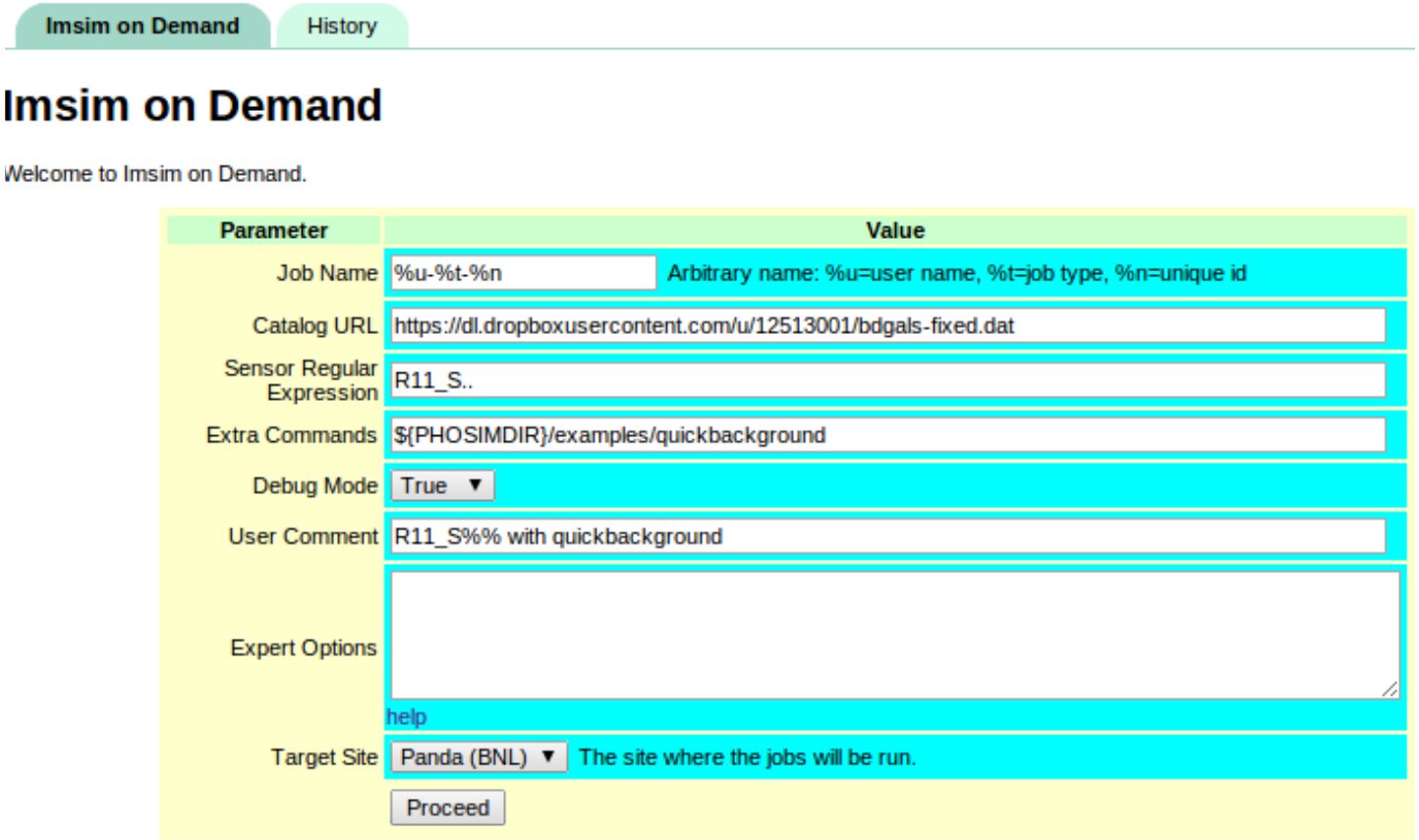
No need for user to obtain grid certificate or learn intricacies of installing tools or running on Grid or at SuperComputing Center

What currently exists



- http://portal.lsstdesc.org/ImsimOnDemand
  - Login with DESC confluence username/password
- Beta-level “Proof of principle”

What currently exists



What currently exists

Imsim on Demand | History

Confirm Imsim on Demand Settings

Messages

Your catalog URL has content length: 95,062,903

Your sensor regular expression matches 9 sensors

Parameter	Value
Job Name	%u-%t-%n
Catalog URL	https://dl.dropboxusercontent.com/u/12513001/bdgals-fixed.dat
Sensor Regular Expression	R11_S..
Extra Commands	${PHOSIMDIR}/examples/quickbackground
Debug	true
User Comment	R11_S%% with quickbackground
Expert Options	
Target Site	Panda (BNL)
Back	Submit

What currently exists

Imsim on Demand | History

Imsim on Demand job submitted

Your job tonyj-ImsimOnDemand-00019 has been submitted.

Your data will be available for download from http://srs.slac.stanford.edu/DataCatalog/?experiment=LSST-DESC&folderPath=/LSST-DESC/ImsimOnDemand/1403121926840

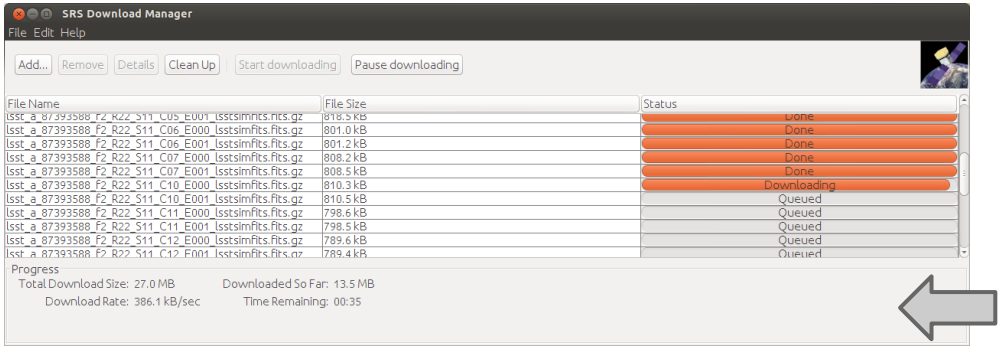
You will be sent an e-mail at tony_johnson@slac.stanford.edu when your job has completed.

You can monitor your job's progress using the Pipeline

```
Dear tonyj,
Job tonyj-ImsimOnDemand-00017 has finished. You can reach your data via:

http://srs.slac.stanford.edu/DataCatalog/?experiment=LSST-DESC&folderPath=/LSST-DESC/ImsimOnDemand/1403113923444

Please use: tonyj@slac.stanford.edu or reply to this e-mail with any concerns/suggestions.
Sincerely Yours,
Imsim On Demand server
```



Folder /LSST-DESC/ImsimOnDemand/1403115818663

Edit description

Name	Type	Files	Events	Size	Created (UTC)	Links
lsst_a_87393588_f2_R22_S11_C00_E000.fits.gz	Dataset	1	0	815.9 kB	18-Jun-2014 18:30:11	Download
lsst_a_87393588_f2_R22_S11_C00_E001.fits.gz	Dataset	1	0	815.5 kB	18-Jun-2014 18:35:28	Download
lsst_a_87393588_f2_R22_S11_C01_E000.fits.gz	Dataset	1	0	808.2 kB	18-Jun-2014 18:30:14	Download
lsst_a_87393588_f2_R22_S11_C01_E001.fits.gz	Dataset	1	0	808.4 kB	18-Jun-2014 18:35:31	Download
lsst_a_87393588_f2_R22_S11_C02_E000.fits.gz	Dataset	1	0	804.1 kB	18-Jun-2014 18:30:17	Download
lsst_a_87393588_f2_R22_S11_C02_E001.fits.gz	Dataset	1	0	804.5 kB	18-Jun-2014 18:35:34	Download
lsst_a_87393588_f2_R22_S11_C03_E000.fits.gz	Dataset	1	0	831.5 kB	18-Jun-2014 18:30:20	Download
lsst_a_87393588_f2_R22_S11_C03_E001.fits.gz	Dataset	1	0	831.3 kB	18-Jun-2014 18:35:37	Download
lsst_a_87393588_f2_R22_S11_C04_E000.fits.gz	Dataset	1	0	811.4 kB	18-Jun-2014 18:30:22	Download
lsst_a_87393588_f2_R22_S11_C04_E001.fits.gz	Dataset	1	0	811.4 kB	18-Jun-2014 18:35:40	Download
lsst_a_87393588_f2_R22_S11_C05_E000.fits.gz	Dataset	1	0	818.4 kB	18-Jun-2014 18:30:25	Download
lsst_a_87393588_f2_R22_S11_C05_E001.fits.gz	Dataset	1	0	818.5 kB	18-Jun-2014 18:35:43	Download
lsst_a_87393588_f2_R22_S11_C06_E000.fits.gz	Dataset	1	0	801.0 kB	18-Jun-2014 18:30:28	Download
lsst_a_87393588_f2_R22_S11_C06_E001.fits.gz	Dataset	1	0	801.2 kB	18-Jun-2014 18:35:46	Download
lsst_a_87393588_f2_R22_S11_C07_E000.fits.gz	Dataset	1	0	808.2 kB	18-Jun-2014 18:30:31	Download
lsst_a_87393588_f2_R22_S11_C07_E001.fits.gz	Dataset	1	0	808.5 kB	18-Jun-2014 18:35:49	Download
lsst_a_87393588_f2_R22_S11_C10_E000.fits.gz	Dataset	1	0	810.3 kB	18-Jun-2014 18:30:34	Download
lsst_a_87393588_f2_R22_S11_C10_E001.fits.gz	Dataset	1	0	810.5 kB	18-Jun-2014 18:35:51	Download
lsst_a_87393588_f2_R22_S11_C11_E000.fits.gz	Dataset	1	0	798.6 kB	18-Jun-2014 18:30:37	Download
lsst_a_87393588_f2_R22_S11_C11_E001.fits.gz	Dataset	1	0	798.5 kB	18-Jun-2014 18:35:54	Download
lsst_a_87393588_f2_R22_S11_C12_E000.fits.gz	Dataset	1	0	789.6 kB	18-Jun-2014 18:30:40	Download
lsst_a_87393588_f2_R22_S11_C12_E001.fits.gz	Dataset	1	0	789.4 kB	18-Jun-2014 18:35:57	Download
lsst_a_87393588_f2_R22_S11_C13_E000.fits.gz	Dataset	1	0	789.2 kB	18-Jun-2014 18:30:43	Download
lsst_a_87393588_f2_R22_S11_C13_E001.fits.gz	Dataset	1	0	788.9 kB	18-Jun-2014 18:36:00	Download
lsst_a_87393588_f2_R22_S11_C14_E000.fits.gz	Dataset	1	0	810.8 kB	18-Jun-2014 18:30:46	Download
lsst_a_87393588_f2_R22_S11_C14_E001.fits.gz	Dataset	1	0	810.3 kB	18-Jun-2014 18:36:03	Download
lsst_a_87393588_f2_R22_S11_C15_E000.fits.gz	Dataset	1	0	795.3 kB	18-Jun-2014 18:30:49	Download
lsst_a_87393588_f2_R22_S11_C15_E001.fits.gz	Dataset	1	0	794.9 kB	18-Jun-2014 18:36:06	Download
lsst_a_87393588_f2_R22_S11_C16_E000.fits.gz	Dataset	1	0	799.7 kB	18-Jun-2014 18:30:51	Download
lsst_a_87393588_f2_R22_S11_C16_E001.fits.gz	Dataset	1	0	800.1 kB	18-Jun-2014 18:36:09	Download
lsst_a_87393588_f2_R22_S11_C17_E000.fits.gz	Dataset	1	0	802.1 kB	18-Jun-2014 18:30:54	Download
lsst_a_87393588_f2_R22_S11_C17_E001.fits.gz	Dataset	1	0	802.3 kB	18-Jun-2014 18:36:12	Download
lsst_e_87393588_f2_R22_S11_E000.fits.gz	Dataset	1	0	953.7 kB	18-Jun-2014 18:30:05	Download
lsst_e_87393588_f2_R22_S11_E001.fits.gz	Dataset	1	0	952.6 kB	18-Jun-2014 18:35:23	Download

List Files . Download Files

What currently exists

Imsim on Demand | **History**

Full History

Note: Clicking on the Status column will take you to the pipeline task that ran the job. Clicking on the Job column will allow you to rerun this task, or a similar one. Clicking on the Output Directory column will take you to the output.

Submit Time	Job	User	Task Type	Status	Output Directory	User Comment
18-Jun-2014 13:09	tonyj-ImsimOnDemand-00019	tonyj	ImsimOnDemand	Running	View dir	R11_S%% with quickbackground
18-Jun-2014 11:52	tonyj-ImsimOnDemand-00018	tonyj	ImsimOnDemand	Terminated	View dir	R11_S%% with quickbackground
18-Jun-2014 11:24	djbard-test-1	djbard	ImsimOnDemand	Success	View dir	does it work at BNL?
18-Jun-2014 11:23	djbard-test-0	djbard	ImsimOnDemand	Success	View dir	does it work at slac?
18-Jun-2014 10:52	tonyj-ImsimOnDemand-00017	tonyj	ImsimOnDemand	Success	View dir	R11_S01 at BNL (Test manual copy)
18-Jun-2014 09:33	tonyj-ImsimOnDemand-00016	tonyj	ImsimOnDemand	Failed	View dir	R11_S01 at BNL (Test manual copy)
18-Jun-2014 04:46	tonyj-ImsimOnDemand-00015	tonyj	ImsimOnDemand	Success	View dir	R11_S01 at BNL (Test file transformation)
17-Jun-2014 17:25	tonyj-ImsimOnDemand-00014	tonyj	ImsimOnDemand	Success	View dir	R12_S01 at BNL
17-Jun-2014 17:22	tonyj-ImsimOnDemand-00013	tonyj	ImsimOnDemand	Success	View dir	R12_S01 at SLAC
17-Jun-2014 16:41	tonyj-ImsimOnDemand-00012	tonyj	ImsimOnDemand	Success	View dir	R12_S01 at SLAC
16-Jun-2014 15:48	jschovan-ImsimOnDemand-00003	jschovan	ImsimOnDemand	Success	View dir	test.jschovan
16-Jun-2014 14:30	jschovan-ImsimOnDemand-00002	jschovan	ImsimOnDemand	Success	View dir	test.jschovan
16-Jun-2014 14:14	tonyj-ImsimOnDemand-00011	tonyj	ImsimOnDemand	Success	View dir	R12_S01 at SLAC

Tools currently used

- SLAC workflow engine
- Web interface and Data Catalog
  - Originally developed for Fermi Gamma-Ray telescope
- Panda distributed analysis system
  - Grid interface (originally used by Atlas)
- Xrootd globally accessible data storage
  - At BNL and SLAC

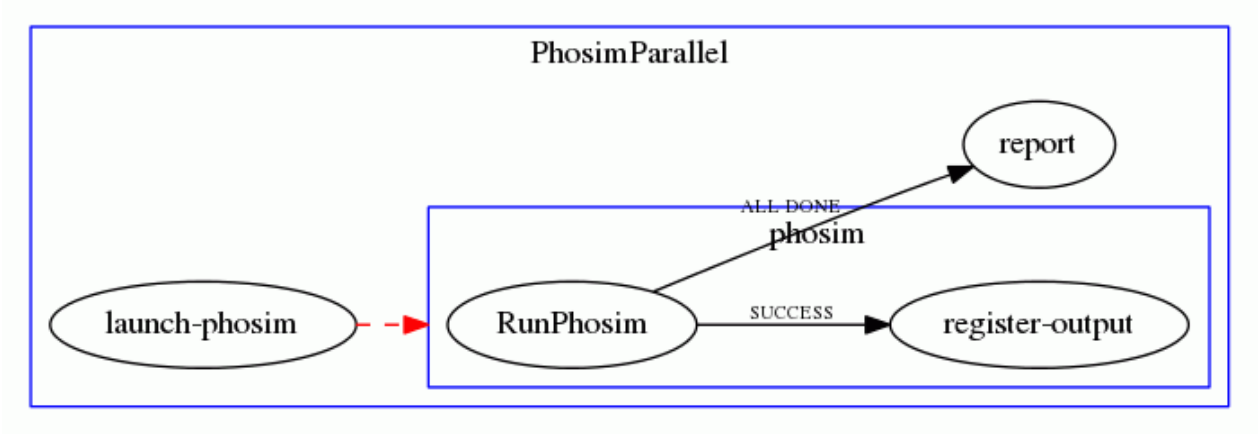
Panda and Grid Tools

- Panda -- Atlas Workload Management and Analysis System
  - Being generalized under ASCR + DOE HEP funded multi-year effort
  - LSST DESC project is early “pilot” of Panda’s use beyond Atlas
  - LSST VO set up -- currently @ BNL, next @ UT Arlington, ...
- CERNVM File System (CVMFS)
  - Allows one central installation of the software to be propagated to Grid sites
  - /cvmfs/oasis.opensciencegrid.org/lsst/test.lsst/bnl/phosim/3.3.2/phosim-3.3.2
- Data Access: xrootd
  - HEP standard distributed data access tool called xrootd (SLAC developed)
  - Exposes data stores to distributed (as well as local) clients, soon via http clients as well as via xrootd tools
  - Also supported by SLAC data catalog and download manager

More Details: https://docs.google.com/presentation/d/1zRez2sUZYTdqFpUaSGPyJviXfO331qXBaM2S1_O6k7s/edit?usp=sharing

Workflow Engine

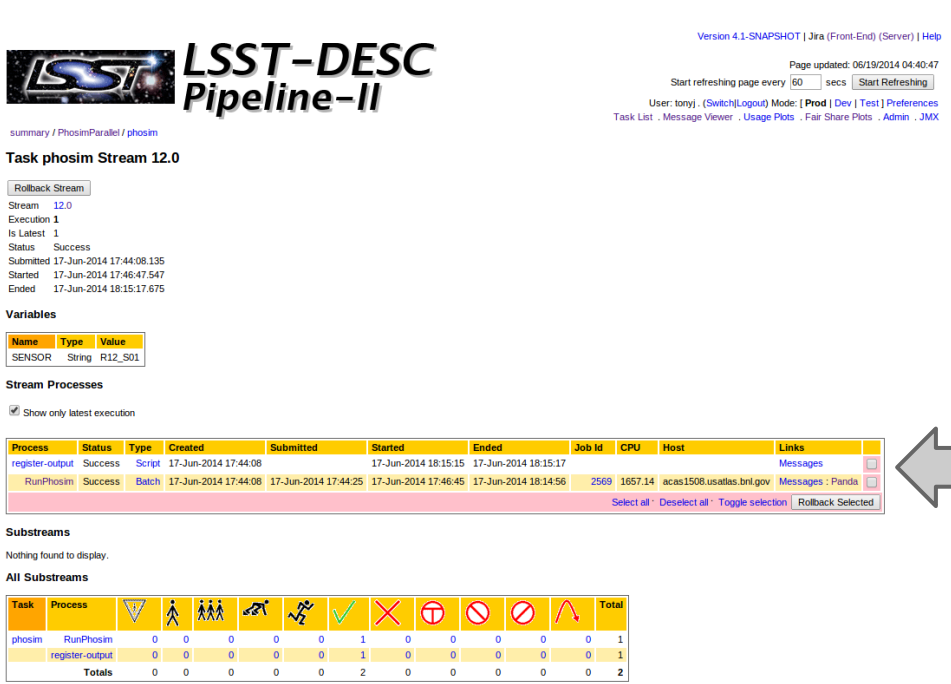
Allows arbitrary workflows to be defined. Can include embedded jython scripts to control flow (for example optional steps) and supports parallelism

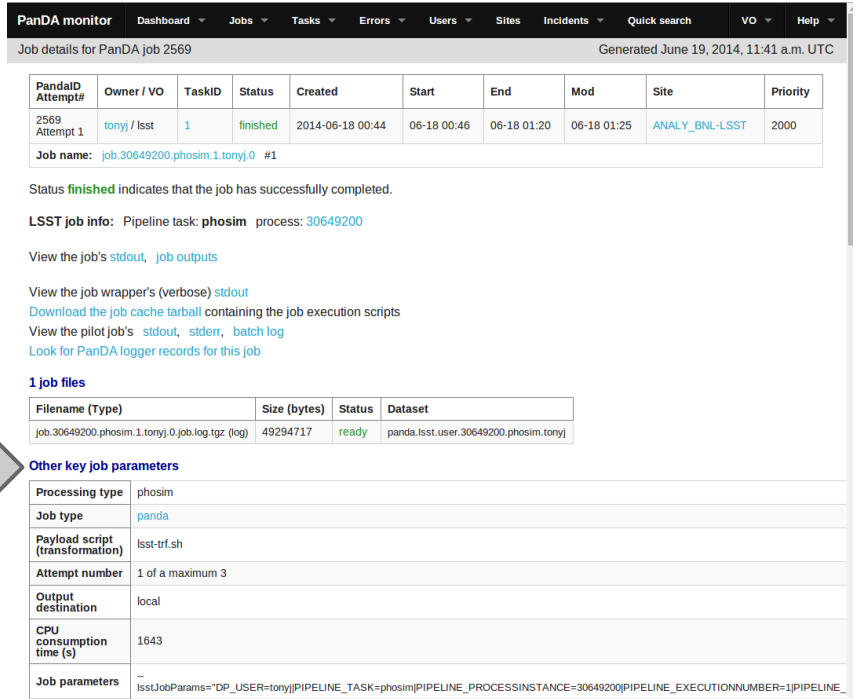


Flexible job submission architecture allows jobs to be submitted to various sites, including SLAC, IN2P3, Grid (Panda, Dirac), and NERSC.

Panda + Workflow engine integration

Workflow engine and Panda support job monitoring, now integrated via links





(Possible) Future Plans

- Incorporate Catsim
- Allow appending multiple catalogs
- More Imsim options (what would be most useful?)
- More interactive web interface (for example select rafts from diagram)
- Ability to view images without downloading
  - For example by using JS9
- Ability to run DMStack tools on images?
  - Producing data as fits, database?
- Alternatives to full phosim?
  - Phenomenological simulation?
- Extend to more than a single image
- Extend to run at more sites
  - Other sites in DESC VO, IN2P3, NERSC
- What would be most useful to you?

Conclusions

- A “Proof of Principle” version of lmsim on Demand now exists
  - Please try it out and give feedback
    - Does it work, does it do what you would like?
- We need input from working groups, individuals on what features would be most useful to add
  - We have fairly regular meetings -- anyone welcome
  - Contact me (tonyj@slac.stanford.edu) to be kept in the loop