



Contribution ID: 361

Type: oral presentation

The GLAST Data Handling Pipeline

Thursday 6 September 2007 16:30 (20 minutes)

The Data Handling Pipeline ("Pipeline") has been developed for the Gamma-Ray Large Area Space Telescope (GLAST) launching at the end of 2007. Its goal is to generically process graphs of dependent tasks, maintaining a full record of its state, history and data products. In cataloging the relationship between data, analysis results, software versions, as well as statistics (memory usage, cpu usage) of the processing it is able to track the complete provenance of all the data products.

The pipeline will be used to automatically process the data down-linked from the satellite and to deliver science products to the GLAST collaboration and the Science Support Center. It is currently used to perform Monte Carlo simulations, and analysis of commissioning data from the instrument. It will be stress tested this summer with "end-to-end" tests of data processing from the satellite and a full 1 year simulation run.

The Pipeline software is written almost entirely in Java and comprises several modules. A set of Java Stored Procedures compiled into the Oracle database allow computations on data to occur without network overhead. The Pipeline Server module accepts user requests, performs remote job scheduling and submission, and processes small "scriptlets" that allow lightweight calculations without the overhead of a batch job. The Pipeline Server submits jobs to the SLAC batch farm (3000+ linux cores), but will soon also submit jobs to a batch farm in France, and via the Grid to a farm in Italy. The "Pipeline Front End" displays live processing statistics via the web. It also provides AIDA charts summarizing CPU and memory usage, average submission wait time and also provides a graphical work-flow representation of the processing logic. Pipeline administrators can interact with the pipeline via web based or line-mode clients.

Submitted on behalf of Collaboration (ex, BaBar, ATLAS)

GLAST

Primary author: FLATH, Dan (SLAC)**Co-authors:** PAVLIN, Igor (SLAC); HEIDENREICH, Karen (SLAC); JOHNSON, Tony (SLAC)**Presenter:** FLATH, Dan (SLAC)**Session Classification:** Distributed data analysis and information management**Track Classification:** Distributed data analysis and information management