

The Distributed Monitoring Framework (DMF)

Brian Tierney, LBNL

The main product of the DMF project is the NetLogger toolkit. NetLogger is a set of libraries and tools to support end-to-end monitoring of distributed applications. During the past few months we have been working closely with the SNfactory project at LBL to help debug and tune their application. According to Stephen Bailey, one of the lead developers on SNfactory project, "NetLogger has been extremely useful in the debugging and commissioning of our data processing pipeline. It has helped us identify bugs and processing bottlenecks in order to improve our efficiency and data quality. It additionally has allowed real time monitoring of the data processing to quickly identify problems that need immediate attention. This debugging, commissioning, and monitoring would have taken much longer without NetLogger."

The Figure below shows a typical workflow for the SNfactory application on a single cluster node. CPU and network data is shown at the bottom. The SNfactory application processes a group of images together, starting with uncompressing the images, and then doing image calibration and subtraction. The next step is to generate a *skyflat* image, which is a calibration image that is formed from a median combine of several of other images. The *skyflat* is used to correct other images to adjust for the sky brightness on a given night, which can vary due to humidity, cloud cover, and so on. The *skyflat* calibration image is then applied to all images within the job.

This figure actually demonstrates a bug in the SNFactory processing that went undetected for several months before NetLogger analysis. Under some conditions it was determined erroneously that the *skyflat* calibration was not necessary. All lifelines except the two nearly vertical ones near the beginning should have converged at the *setskyflat* event.

In addition, the DMF project members are working with Harvey Newman's group at Caltech to deploy NetLogger in CMS and integrate it with their MonaLisa tool. NetLogger is now being used in production CMS jobs to track job progress. NetLogger lifelines provide an intuitive mechanism for exploratory visual analysis of Grid applications. Dr Newman states "I believe that the ability to track and troubleshoot Grid operations, from the applications to the Grid middleware, to site and network performance, will be critical if we are to attain the vision of globally scalable Grids able to serve the needs of the largest scientific projects in HEP as well as many other fields of science. In my opinion, the NetLogger approach is very well designed to meet these needs, and will be a key factor allowing this to happen."

NetLogger has also been incorporated in to a number of Grid middleware projects, such as pyGlobus, pyGridWare, and Akenti. In August 2004 we released a new version of NetLogger, called "NetLogger-lite", that only contains NetLogger's most commonly used features and is easier to install and use.

In addition to our NetLogger work, The DMF project members have been integral to two GGF documents. Our co-leading of the Discovery and Monitoring Event Descriptions (DAMED) working group, helped produce a description of how the attributes of monitoring data should be represented in a consistent and straightforward way (See: <http://www.ggf.org/documents/GWD-I-E/GFD-I.025.pdf>). As an example of this

standard, in the NetLogger Toolkit we provided an efficient wire representation that maps to the DAMED data model. We also co-lead the Network Measurements working group, which produced a standards document describing a set of network characteristics that are useful for Grid applications and services as well as a classification hierarchy for these characteristics. (see: <http://www.ggf.org/documents/GWD-R/GFD-R.023.pdf>) The continuation of this work has led to interoperability schemas that are being used by Internet-2 (<http://www.internet2.edu/>) and Dante (<http://www.dante.net/>) to publish network monitoring data.

