20th International Conference on Computing in High Energy and Nuclear Physics (CHEP2013)

Contribution ID: **391** Type: **Poster presentation**

SPADE : A peer-to-peer data movement and warehousing orchestration

Monday 14 October 2013 15:00 (45 minutes)

The SPADE application was first used by the IceCube experiment to move its data files from the South Pole to Wisconsin. Since then is has been adapted by the DayaBay experiment to move its data files from its experiment, just outside Hong Kong, to both Beijing an LBNL. The aim of this software is to automate much of the data movement and warehousing that is often done by hand or home-grown script on smaller experiments.

The latest generation of this software has been developed to be experiment independent and fully peer-to-peer. This means that it can be used anywhere, not only to move raw experiment data to computing centers but also as a means of exchanging data between centers. For example, gathering Monte Carlo production data from member institutions to an experiment's main warehouse and distributing copies back out to the users who need them.

Special attention has been paid to keeping administration and infrastructure time and effort to a minimum so that this software can be easily used by experiments that do not have large computer support resources. Additional features of the software include facilities to run prompt analysis on files as they arrive in the warehouse, the archiving of files to HPSS and the option to create local copies of files before they are moved to enable immediate assessment of the data by computers based at the experiment.

This paper will examine the requirement of the new version of the software. It will then discuss the main architecture of the application and show how this satisfies those requirements. This will be followed by a quick tour of what is needed to install and run the basic version, after which an explanation of now experiment specific customization can be made.

Finally, the DayaBay use case will be summarized to show a practical deployment of the application.

Author: Dr PATTON, Simon (LAWRENCE BERKELEY NATIONAL LABORATORY)

Co-author: Ms MACKENZIE, Cindy (University of Wisconsin - Madison)

Presenter: Dr PATTON, Simon (LAWRENCE BERKELEY NATIONAL LABORATORY)

Session Classification: Poster presentations

Track Classification: Distributed Processing and Data Handling A: Infrastructure, Sites, and Virtualization