



Contribution ID: 16

Type: Poster

## Scientific Cluster Deployment & Recovery: Using puppet to simplify cluster management

*Tuesday, May 22, 2012 1:30 PM (4h 45m)*

Deployment, maintenance and recovery of a scientific cluster, which has complex, specialized services, can be a time consuming task requiring the assistance of Linux system administrators, network engineers as well as domain experts. Universities and small institutions that have a part-time FTE with limited knowledge of the administration of such clusters can be strained by such maintenance tasks.

This current work is the result of an effort to maintain a data analysis cluster with minimal effort by a local system administrator. The realized benefit is the scientist, who is the local system administrator, is able to focus on the data analysis instead of the intricacies of managing a cluster. Our work provides a cluster deployment and recovery process based on the puppet configuration engine allowing a part-time FTE to easily deploy and recover entire clusters with minimal effort.

Puppet is a configuration management system (CMS) used widely in computing centers for the automatic management of resources. Domain experts use Puppet's declarative language to define reusable modules for service configuration and deployment.

Our deployment process has three actors: domain experts, a cluster designer and a cluster manager. The domain experts first write the puppet modules for the cluster services. A cluster designer would then define a cluster. This includes the creation of cluster roles, mapping the services to those roles and determining the relationships between the services. Finally, a cluster manager would acquire the resources (machines, networking), enter the cluster input parameters (hostnames, IP addresses) and automatically generate deployment scripts used by puppet to configure it to act as a designated role. In the event of a machine failure, the originally generated deployment scripts along with puppet can be used to easily reconfigure a new machine.

The cluster definition produced in our cluster deployment process is an integral part of automating cluster deployment in a cloud environment. Our future cloud efforts will further build on this work.

**Primary author:** HENDRIX, Valerie (Lawrence Berkeley National Lab. (US))

**Co-authors:** BENJAMIN, Doug (Duke University (US)); Mr YAO, Yushu (LBNL)

**Presenter:** HENDRIX, Valerie (Lawrence Berkeley National Lab. (US))

**Session Classification:** Poster Session

**Track Classification:** Computer Facilities, Production Grids and Networking (track 4)