

Grid Systems Deployment & Management Using Rocks

Karan Bhatia, Sandeep Chandra, Federico Sacerdoti (San Diego Supercomputer Center)

Objective: Design a centralized mechanism to define and roll out the software stack to hardware resources at the partner institutions in GEON grid.

Description: Computational clusters have become the dominant platform for a wide range of scientific disciplines. Cluster management software is available that allows administrators to specify and install a common software stack on all cluster nodes, and enable centralized control and diagnostics of components with minimal effort. While grid deployments have similar management requirements, they have faced a lack of available tools to address their needs. These grids seek to offer a common operating environment for a community or scientific domain, and typically involve a diverse set of resources operating in a geographically dispersed environment. We discuss grid design from the perspective of GEON, although its similarity to other grid efforts makes our results applicable to other projects.

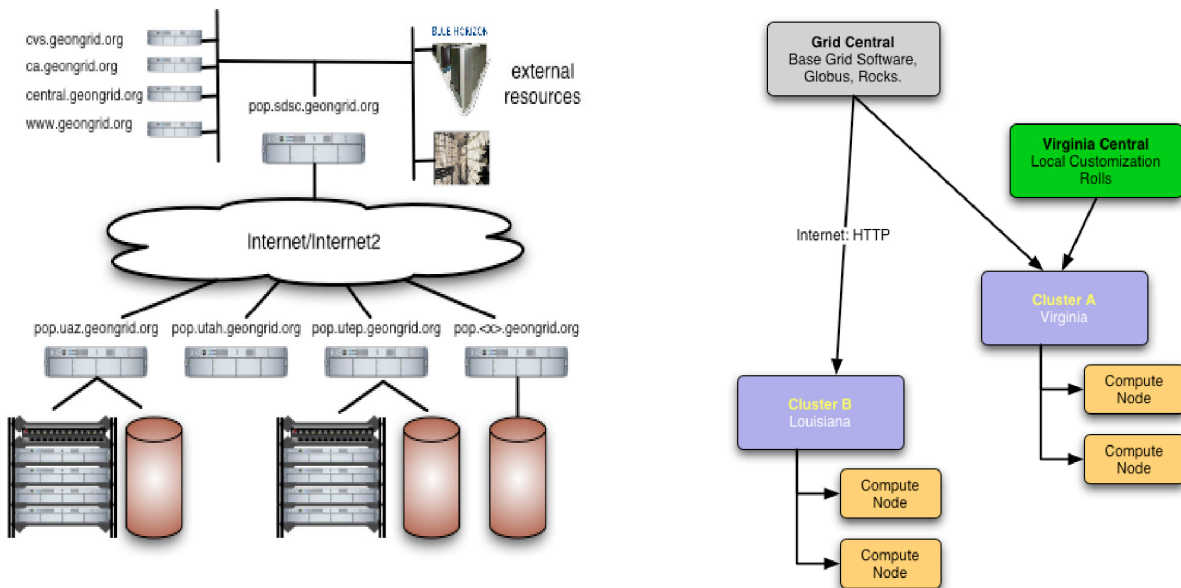


Figure 1: GEONGrid Wide Area Deployment Architecture using Rocks

Figure 1 shows the GEON grid architecture. A set of physically distributed clusters are located within the administrative domain of each of many participating sites. These clusters may have zero or more compute nodes and may consist of different hardware architectures. The operation of this virtual organization requires the machines to run a common software stack that enables interoperability with other GEON resources. In addition to being geographically distributed, the clusters at each partner site build upon the common software base to provide site-specific applications and services. The challenge, therefore, is to manage the distributed set of hardware resources, physically distributed at partner institutions and connected over the commodity

Internet, in a way that minimizes system administration costs while achieving interoperability and local site autonomy. We use the Rocks cluster distribution as a starting point for the GEON effort. With our direction, the Rocks toolkit has been modified to support wide-area deployment and management of systems in a fashion appropriate for GEON like grids.

The wide area grid deployment involves a Central server that holds the software stack. Frontend pop nodes of the grid obtain a full Rocks cluster distribution from the central; this distribution is suitable to install local compute, data, and customized nodes. We require a small static bootstrap environment for the pop initialization, which contains only network drivers and hardware-probing features. This bootstrap environment is stored on the node for use in the upgrade process. Figure 1 shows the wide-area kickstart architecture. The central server holds the base Rocks distribution, including the Linux kernel and installation environment, along with standard rolls such as the HPC (high performance computing), and the domain-specific GEON roll. We initialize GEON frontend pops over the wide-area Internet from this central. The software disbursement methodology is optimized for unreliable networks. The frontend pop locally integrates the Rocks base and the various Rolls selected from the central server, and the combined software stack is rebound into a new Linux distribution. The pop then initializes and installs the cluster nodes with its combined distribution.

Current Status: To address the primary requirement of the GEON grid, a low management cost for grid-wide software installation and updates, we extended the Rocks distribution to perform full cluster installations over wide area networks. While compute nodes in Rocks always employed the LAN to install software from the frontend machine, the frontend itself had to be integrated with CD or DVD media. This strategy, while appropriate for cluster instantiations of basic Rocks software, is insufficiently flexible for the dynamic needs of the GEON grid. Specifically, we considered affecting grid-wide software with mailed disk media unacceptable.

This functionality has since been rolled into the base distribution of Rocks and Rocks hosts a central server at SDSC, to serve the rest of the cluster and grid community, that supports full cluster installation over wide area networks.

References:

- K. Bhatia, S. Chandra, F. Sacerdoti, Grid Systems Deployment and Management using Rocks, IEEE Cluster 2004, San Diego, CA, Sep 2004.
- NPACI Rocks: Tools and Techniques for Easily Deploying Manageable Linux Clusters, Philip M. Papadopoulos, Mason J. Katz, and Greg Bruno, Cluster 2001.
- Thomas E. Anderson, David E. Culler, and David A. Patterson. A case for NOW (networks of workstations). IEEE Micro, 15(1): 54–64, February 1995.
- B. E. Finley. VA SystemImager. In Proceedings of the 4th Annual Linux Showcase and Conference, Atlanta, GA, Oct. 2000.
- Scyld Software. Scyld Beowulf Clustering for High Performance Computing.
- Yellow dog Updater, Modifies <http://linux.duke.edu/projects/yum/>
- NSF Middleware Initiative <http://www.nsf-middleware.org/>