Use of the Adaptable Profile-driven Testbed (Apt) for Dynamically Provisioning HPC Resources

Brian Haymore, Joseph Breen, Anita Orendt, Thomas Cheatham, Julio Facelli, Steven Corbato, Center for High Performance Computing, University of Utah

Last autumn, NSF funded a collaboration led by Rob Ricci of the School of Computing's Flux group, in conjunction with the Center for High Performance Computing (CHPC) at the University of Utah to develop an "adaptable profile-driven testbed". This hardware testbed is designed to allow computational science and computer science research teams to use the same physical resources for different missions, e. g., network experiments, high performance computing experiments, security experiments, calculations requiring HIPPA compliant environments, etc. The Flux group's emphasis is to create a low entry barrier for creating very reproducible experiments. CHPC's emphasis is to create environments to roll out HPC images on demand, to scale the images dynamically as additional resources become either available or current resources become unavailable (i.e., provisioned to a different task or experiment), and to support multiple images with different HPC and security contexts.

Apt starts with technology and lessons learned from the previous Flux testbeds and expands the scope. By leveraging specific hardware and by expanding the software, Apt is able to provide an environment for researchers in the traditional network research community, the HPC community, as well as other communities through Apt's "on demand" profiles. Experimenters can create profiles, save them, and re-use them to repeat experiments, or, to share with other researchers.

The Apt cluster contains two classes of nodes:

- 128 Dell PowerEdge r320 nodes, each with a single Intel Xeon E5-2450 processor (8 cores, 2.1Ghz), 16GB Memory, and four 500GB Hard Drives
- 64 Dell PowerEdge c6220 nodes, each with dual Intel Xeon E5-2650v2 processors (8 cores, 2.6Ghz), for a total of 16 cores, 64GB Memory, and two 1TB Hard Drives

These nodes connect via a 1Gbps Ethernet control network to the network switch that also has multiple high bandwidth data plane network options, including 10Gpbs, 40Gbps,and 56Gbps Ethernet and also FDR Infiniband. Additional detail on the hardware is available at http://docs.aptlab.net/hardware.html.

Flux team members are developing software to dynamically provision this hardware to meet the needs of the researchers. The users define a profile which includes all the information needed to run an experiment. This profile includes the description of the resources, both hardware and software, needed for the experiment, and provides the mechanism to enable repeatable and reproducible research.

CHPC is exploring ways to leverage the dynamic nature of the Apt testbed for use by the HPC community as a cloud resource. In this effort, we have started with our existing HPC cluster environment as the basis to establish a traditional HPC profile on the Apt hardware as a cluster called Tangent, to launch jobs on the C6220 nodes of the Apt cluster, using the cloud support features of SLURM for the scheduling of jobs. This profile will mount current CHPC file systems and have CHPC applications

available. From the user perspective, access to this resource is obtained via a login to an interactive node for the Tangent cluster. The Tangent interactive nodes are local to CHPC and allow users the users to submit batch jobs that will spin up dynamic HPC images on the Apt hardware.

For managing the jobs, CHPC is developing a workflow to provide the status of the individual nodes to the scheduler and to create an Apt experiment to provision the resources as a Tangent cluster compute node. When a user submits a job, a check if the requested resources are available on the Apt cluster is made. If the resources on the Apt cluster are not available, the job will sit idle in the batch queue until the next cycle of the script. When the resources are available, the script will create an experiment using CHPC's HPC profile and provision the resources for the requested wall time. At this stage, the user will see that the system is allocating the node(s) for the job. Once provisioned, the script launches the job. The job will run until complete or until the job reaches the requested wall time. Once complete, the script marks the node(s) as unavailable, de-provisions them from the Tangent cluster environment, and returns them to the Apt resources for use by another experiment.

Currently we have only a minimal scheduling system, but we are working towards more complete integration between the tangent batch queue and the Apt hardware, as well as having additional capabilities in the scheduling, such as dealing with reservations to deal with special requests.

CHPC is currently testing several scientific applications using this system in order to learn more about the overhead for the provisioning and de-provisioning of nodes in this dynamic manner. In addition, we are inviting select CHPC users to test the system, in order to provide feedback to improve the user experience. This will also provide us with a method to monitor interaction between the different usage cases that co-exist on the hardware. We are also starting to develop a compliance regulated (e.g., HIPAA) HPC image based on the protected environment used at CHPC.