Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research

Kate Keahey
Argonne National Laboratory

Cloud computing services have become critical to all major 21st century economic activities – yet, we are only beginning to understand this new important paradigm. Questions persist regarding applicability of the cloud platform to the emergent data-intensive and sensor-based applications, its suitability for high performance computing (HPC) applications, and its potential to leverage major emergent technologies such as Software Defined Networking (SDN). Answering those questions requires the ability to perform experiments at scale – in other words, an experimental testbed that would support experimentation with Big Data, Big Compute, and Big Instrument problems.

The Chameleon project will provide such a large-scale platform to the open research community, allowing them to explore transformative concepts in deeply programmable cloud services, design, and core technologies. It's reconfigurability will allow users to explore problems ranging from the creation of Software as a Service to kernel support for virtualization. The broad range of supported research will include areas such as developing Platforms as a Service, creating new and optimizing existing Infrastructure as a Service components, investigating software-defined networking, and optimizing virtualization technologies. While the hardware, and the general capabilities of the Chameleon experimental testbed have been broadly defined, the details of the approach are still open to feedback from users wanting to run particular experiments.

The Chameleon testbed, deployed at the University of Chicago (UC) and the Texas Advanced Computing Center (TACC), will consist of 650 multi-core cloud nodes (~14,500 cores total), over 5PB of total disk space, and leverage 100 Gbps connection between the sites. A large part of the testbed will consist of homogenous hardware to support large-scale experiments. This part of the testbed is composed of 12 racks, comprising 46 Xeon Haswell processors (42 compute and 4 storage servers) with OpenFlow-enabled switches; each rack will have 128 TB of storage and one of them will contain Infiniband network. In addition to distributed storage nodes, Chameleon will have 3.6 PB of central storage, for a persistent object store and shared filesystems. The testbed will also support heterogeneous units consisting of Atom microservers, ARM microservers, as well as a mix of servers with high RAM, FPGAs (Xilinx/Convey Wolverine), NVidia K40 GPUs, and Intel Xeon Phis to allow experimentation with high-memory, large-disk, low-power, GPU, and co-processor units. In its initial phase, the project leverages existing FutureGrid hardware at the University of Chicago and the Texas Advanced Computing Center with their FutureGrid configuration (i.e., as OpenStack clouds) to provide a transition period for the existing FutureGrid community of experimental users. This part of the testbed, called FutureGrid@Chameleon, is available now.

To support a broad range of experiments, Chameleon will support a graduated configuration system allowing full user configurability of the software stack, from provisioning of bare metal and network interconnects to delivery of fully functioning cloud environments. In addition, to facilitate experiments and provide a "one stop shopping" for experimental artifacts, Chameleon will support a set of services designed to meet researchers needs, including support for experimental management, reproducibility, and repositories of trace and workload data of production cloud workloads based on both commercial and scientific clouds. The project will also provide innovative ways of integrating testbeds into the educational pipeline by designing and publishing new educational artifacts such as ready to deploy Chameleon appliances.

The project is led by the Computation Institute at the University of Chicago and partners from the Texas Advanced Computing Center at the University of Texas at Austin, the International Center for Advanced Internet Research at Northwestern University, the Ohio State University, and University of Texas at San Antonio, comprising a highly qualified and experienced team. The team includes members from the NSF supported FutureGrid project and from the GENI community, both forerunners of the NSFCloud solicitation under which this project is funded. Chameleon will also form a set of partnerships with commercial and academic clouds, such as Rackspace, CERN and Open Science Data Cloud (OSDC), and will partner with other testbeds, notably GENI and INRIA's Grid'5000 testbed.