Chameleon: Cloud computing for computer science

26 August 2014

The color-changing lizard is a fitting mascot for the project, which will create a customizable, large-scale cloud computing facility for cloud computing research. Credit: Matt Stelmaszek, Texas Advanced Computing Center (TACC)

Cloud computing has changed the way we work, the way we communicate online, even the way we relax at night with a movie. But even as "the cloud" starts to cross over into popular parlance, the full potential of the technology to directly impact science, medicine, transportation, and other industries has yet to be realized.

To help investigate and develop this promising cloud computing future, the National Science Foundation (NSF) has announced a new $10 million project to create a cloud computing testbed called Chameleon, an experimental testbed for cloud architecture and applications. This testbed will enable the academic research community to develop and experiment with novel cloud architectures and pursue new, architecturally-enabled applications of cloud computing, specifically for the computer science domain.

Cloud computing refers to the practice of using a network of remote servers to store, manage and process data, rather than a local server or a personal computer. In recent years, cloud computing has become the dominant method of providing computing infrastructure for Internet services. While most of the original concepts for cloud computing came from the academic research community, as clouds grew in popularity, industry has driven their design and development.

"Like its namesake, the Chameleon testbed will be able to adapt itself to a wide range of experimental needs, from 'bare metal' reconfiguration to support for ready-made clouds", said Kate Keahey, a scientist at the Computation Institute at the University of Chicago and principal investigator for Chameleon.

Chameleon is designed to support a variety of cloud research. To support users building cloud services and platforms, Chameleon will include persistent infrastructure clouds. To support researchers investigating low-level software for clouds, Chameleon will provide "bare metal" provisioning of hardware where users specify, and can modify, the full software stack they will experiment on. For researchers that want dedicated, but not fully custom environments, Chameleon will provide pre-configured software stacks that are provisioned on bare metal.

The lead institution on the project is the Computation Institute (CI) at the University of Chicago, along with partners at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin, the International Center for Advanced Internet Research at Northwestern University (iCAIR), Network-Based Computing Laboratory at The Ohio State University, and the UTSA Cloud and BigData Laboratory at The University of Texas at San Antonio. This highly qualified and experienced team includes members from Rackspace, as well as the NSF-supported FutureGrid project and from the Global Environment for Network Innovations (GENI) community, both forerunners of the NSF CISE Research Infrastructure: Mid-Scale Infrastructure – NSFCloud program that funds the Chameleon project. Technology partners on the project are Dell Inc. and Intel.

TACC Executive Director Dan Stanzione, said: "This is UC and TACC's first machine targeted specifically at cloud and computer science research, and it will support computer science departments across the country. It gives us a platform to study the problem of cloud computing and to do computer science research that can't be done on commercial clouds or on production supercomputing systems."

Although Chameleon will have a different focus than TACC's other production systems like Stampede and Lonestar, which enable computational science, it will provide a learning platform for how to build better large-scale computer systems. Computer science faculty members, cloud researchers, and others will benefit from this new testbed.

"Finding suitable realistic experimental platforms to facilitate computer science research is always a daunting task, and that includes research in cloud computing," said Patricia Teller, a renowned professor in Computer Science at The University of Texas at El Paso. "In some cloud environments job scheduling must consider information regarding node resources, e.g., processing, storage, and power reserves, and effective mapping of applications to node architectures is essential."

"With the assistance of the FutureGrid project team, our students implemented a tactical cloudlet in our lab. Although this got us started with our experimentation, it is not sufficient for our investigation of new Mobile Ad hoc Network (MANet) communication protocols, scheduling policies for cloudlets and clouds, and various cloud provisioning algorithms that can be used for clouds comprised of multiple cloudlets. Our students need to experiment at scale, with hardware platforms to which they do not have access, and with multiple cloudlets within a cloud. They cannot wait to get access to Chameleon!"

"We're going to let users become system administrators on Chameleon," said Warren Smith, a TACC research associate in web and cloud services, and co-principal investigator on the FutureGrid project. "We can isolate parts of the machine and let users perform large-scale experiments where they want full control of the software environment from the operating system up."

The Chameleon testbed will consist of 650 cloud nodes with five petabytes of storage. Researchers will be able to mix-and-match hardware, software and networking components and then test their performance. This flexibility is expected to benefit many scientific communities, including the growing field of cyber-physical systems, which integrates computation into physical infrastructure to test the efficiency and usability of different cloud architectures on a range of problems, from machine learning and adaptive operating systems to climate simulations and flood prediction. The research team plans to add new capabilities in response to community demand or when innovative new products are released.

Another aspect that makes Chameleon unique is its support for heterogeneous architectures, including low power processors, general processing units (GPUs) and field programmable gate arrays (FPGAs), as well as a variety of network interconnects and storage devices.

A New Pipeline of Researchers

In addition, Stanzione says, this is a chance for TACC to work closely with UTSA, which is one of the 15 University of Texas System institutions, and Rackspace, a leader in hybrid cloud and founder of OpenStack®, the open-source operating system for the cloud.

"We welcome the opportunity to work with UTSA and the other Chameleon partners to build a stronger pipeline with their education and outreach community." Led by UTSA with the involvement of UTSA students, TACC will provide novel educational materials and training activities that include tutorials at major conferences, summer schools, and online courses.

TACC has been a long time provider of advanced computing and visualization resources as well as support services to enable researchers to learn to use advanced computing effectively. TACC's staff members strive to provide the technology,

computing, and visualization assistance that empower UT System's computational research community to impact science and change the world.

FutureGrid Foundation and Chameleon Timing

Chameleon will originally rely on the resources of the FutureGrid project at the CI and TACC. In the fall of 2014, those resources will be used to provide a seamless transition for FutureGrid users. By the spring of 2014, these resources will support "bare metal" provisioning and will eventually be supplanted by the first new hardware purchase in Summer/Fall of 2015.

"In a project like this, you would typically have to wait for roughly a year as the new hardware is built," said Keahey. "Leveraging FutureGrid resources allows us to hit the ground running and start serving the community right away while also providing transition for FutureGrid users."

Chameleon is joined by a second project, Cloud Lab, as the recipient of grants from the NSFCloud program, which hopes to propel cloud computing technology into its next era. Ultimately, the goal of the NSFCloud program and the two new projects is to advance the field of cloud computing. These awards are the first step in meeting this goal.

"Just as NSFNet laid some of the foundations for the current Internet, we expect that the NSFCloud program will revolutionize the science and engineering for cloud computing," said Suzi Iacono, acting head of NSF's Directorate for Computer and Information Science and Engineering. "We are proud to announce support for these two new projects, which build upon existing NSF investments in the Global Environment for Network Innovations (GENI) testbed and promise to provide unique and compelling research opportunities that would otherwise not be available to the academic community."

Provided by University of Texas at Austin