Tapis computing platform weaves together science computing tools

October 23 2019, by Jorge Salazar

Development has started on the Tapis software platform, which aims to help researchers more easily leverage powerful supercomputers and integrate and manage data from different and distant sources. Credit: TACC/UH

Scientists looking to reduce their complexity to research and add a new computational tool to their tool belt can explore the Tapis Project. The Tapis software platform aims to help researchers more easily leverage powerful supercomputers and integrate and manage data from different and distant sources.

The National Science Foundation (NSF) awarded a $2.9 million grant to the Texas Advanced Computing Center (TACC) and The University of Texas at Austin (UT Austin), in addition to a $1 million award to the University of Hawaii (UH). The NSF awards started September of 2019 and supports continued development of Tapis, short for TACC-APIs and plays off of the word tapestry—weaving together services and capabilities. An application programming interface (API) is an interface to a software system that has been built or engineered for another program to use.

"Tapis is a research computing platform for computational science and computational research," said principal investigator (PI) Joe Stubbs, who manages the Cloud and Interactive Computing Group at TACC. "Tapis is a software system that helps researchers use the supercomputers and other kinds of computing resources that we have here at TACC and at other places."

"The easiest way to describe Tapis is that it's a web-based application that provides all the tools a modern scientist needs to do data-intensive, computationally-intensive research," said Co-PI Gwen A. Jacobs, Director of Cyberinfrastructure, University of Hawai'i System. "One of the things that's different about Tapis is that it weaves together all the important tools that the researcher needs. That's the real power of Tapis."

Tapis will serve a diverse group of users with varying expertise in using computational tools for their research. On one end of the spectrum will be 'power users' with extensive experience of advanced computing resources and programming. Tapis will help them automate and streamline their large workflows or pipelines of software applications.

On the opposite end of the spectrum are scientists just beginning to tap into the possibilities of applying advanced computing to their research.

"What we're trying to do for them with Tapis," said Stubbs, "is have the easiest road to entry on running computational programs on the supercomputers."

And then there's the group in the middle, typically large software development projects focused on specific research domains, such as immunology, astronomy, or bioinformatics.

"The goal with Tapis is to enable researchers to access these computational resources in a more user-friendly way," said Stubbs.

The NSF-funded computational resources are broadly described as cyberinfrastructure, the online ecosystem shared by researchers, backed up by advanced computing resources, hosted in data centers, and supported by experts. "Web developer teams and other developers on those cyberinfrastructure projects can leverage Tapis to build their cyberinfrastructure project more quickly."

To these ends, one example is the TAPIS API framework support for streaming sensor data, where in a complex workflow, one event, such as a detection on a sensor array, can trigger another event, and so on, or even multiple analysis routines.

"Event-driven computing," explained Jacobs, "means that the workflow isn't running all the time. That's a great feature for scientists who have to acquire their data sporadically, where they're getting data from sources such as sensors and data uploads. This means that they don't have to run all the code manually. Once the workflow is set up, it can be hands-free computing, in a way, hands-free analysis."

Tapis will integrate the Cloud-Hosted Real-time Data Services for the Geosciences (CHORDS) project, part of the NSF-funded EarthCube, to achieve event-driven computing.

The APIs applied to science allow different systems to talk to each other, in a sense. "The idea with Tapis," said Stubbs, "is to have a machine-readable and consumable interface to computational resources, like supercomputers, but also high performance storage systems, like our Corral storage system, or our global file system, Stockyard, and other filesystems across the country. We want to have an interface that is easily accessed and manipulated in other programs."

Another feature Tapis will offer is a new security kernel, which acts like a gate that controls access to system resources. The Tapis security kernel will be decentralized, allowing scientists to more easily stand up their own applications and retain local control over confidential data.

"The new security kernel allows us to offer all the managed security, authentication, and authorizations that have been done in the past," said Co-PI Sean Cleveland, a cyberinfrastructure research scientist at the University of Hawaii. "But It will also allow data centers and institutions to deploy their own security kernel, so they can use their own user credentials and manage their own security in their own way, as well as deploy individual components of the framework at their institution, and be able to leverage some of the centralized work. It's a new, hybrid system of using the science-as-a-service, platform-as-a-service, but if you want more control and customization, you can deploy smaller pieces on site and still be able to leverage some of the larger, managed components for different needs."

Tapis will give users the ability to simplify the process of creating applications, a powerful tool for scientists. "If you can program a workflow and have that workflow run in a platform like Tapis, that makes the process easier because all of the components can talk to each other more easily," said Jacobs. "That means that the investigator has to construct that workflow once. Then they save that workflow as an application within the Tapis infrastructure and reuse it."

Saving all the parameters of the software environment will also enable scientists to go back and run the data analysis again at a later date, which promotes scientific reproducibility.

"This really is a complete collaboration between TACC and the University of Hawaii," explained Stubbs.

TACC brings extensive expertise in high performance computing and in building distributed software systems. The components of Tapis themselves can run on commodity, or off-the-shelf servers, although some components at TACC will run on the NSF-funded Jetstream cloud.

Team members at UH are contributing to the development, design, and architecture of the Tapis system. What's more, they bring access to an abundance of important domain research unique to Hawaii in areas such as climate, ocean, coral reefs, human microbiome, and population studies around health disparities.

"Having the Tapis project for us here in Hawaii is a huge awareness boost for applying advanced cyberinfrastructure to data intensive science," said Jacobs. "Without a project like this, many of our investigators might not be aware of these resources."

One of the major milestones the investigators are working toward is an end-of-year workshop for early adopters in the summer of 2020. "The idea is to have the workshop where we invite the researchers to come, bring their data sets, to give presentations on their science and use case, but also for the Tapis team to present on the capabilities of the system by the end of year one," said Stubbs.

"We are really excited to launch the new NSF-funded Tapis project," said Co-PI Maytal Dahan, Director of Advanced Computing Interfaces

at TACC. "Tapis will transform scientists' productivity by facilitating the discovery, access and use of powerful cyberinfrastructure capabilities and services. We want to reduce the complexity to accomplish science and improve the time-to-science by offering a variety of secure and robust API services that can support our users in a production-quality environment.

The TACC team will work on various aspects of the project—development of a security kernel, streaming data APIs and integration, quality assurance and continuous integration testing, outreach, training and workforce development. I am really proud of the team, both at TACC and UH, and we are all enthusiastic to work together with the scientific community from the onset via our early adopters' program to create services that make a positive impact on the scientific community."

The Tapis project is funded as part of the Cyberinfrastructure for Sustained Innovation (CSSI), a crosscutting NSF program lead by the Office of Advanced Cyberinfrastructure (OAC). "CSSI supports the development of innovative cyberinfrastructure that enables communities of researchers to continue and accelerate advances in all fundamental science and engineering domains supported by NSF," said Dr. Stefan Robila, the Program Director in OAC that manages the award. "By building on prior work and leveraging existing leadership computational resources such as those available at TACC, Tapis contributes to continuous strengthening of the national cyberinfrastructure, while at the same time lowering the barriers in accessing it."

Provided by University of Texas at Austin