Python bindings snake into global arrays toolkit

September 26th, 2011 in Technology / Computer Sciences

Researchers at DOE's Pacific Northwest National Laboratory have expanded the Global Arrays Toolkit to support the computer programming language Python to make it easier for programmers to write codes and incorporate features of GA.

While many of us don't want anything to do with snakes, for some, a certain kind of Python—the computer programming language, that is—is the preferred option. Researchers at Pacific Northwest National Laboratory have expanded the Global Arrays (GA) Toolkit to include full support for Python and make it easier for programmers to write codes and take advantage of GA features.

Use of Python in the high-performance computing community is growing, with multiple research institutions—including national laboratories as well as the Ohio Supercomputing Center—turning to the easy-to-use language to write high-performance computing codes. Python is used in scientific applications, such as bioinformatics, visual analytics, molecular dynamics, hydrology, material sciences and more.

The GA Toolkit enables researchers to more efficiently access global data, run bigger models, and simulate larger systems, resulting in a better understanding of the data and processes being evaluated. Integrating Python with GA allows programmers to more easily customize the GA Toolkit when they need shared memory for a distributed memory computers, improving and expanding researchers' ability to access the necessary data.

The Python bindings for GA were developed using the Cython language, which makes writing C extensions for the Python language as easy as using Python itself. The GA Python bindings provide the programmer with access to local memory or copies of remote memory as NumPy arrays, allowing full use of the rich NumPy application programming interface (API) to extend the functionality of Global Arrays beyond previous capabilities. This combination of GA and NumPy was then used to create a distributed, work-alike replacement of the NumPy module.

By integrating Python with GA, programmers are provided with a more convenient globally-shared view of multi-dimensional arrays while retaining the option to use the Message Passing Interface (MPI) if needed. Using GA with Python is significantly simpler than MPI for many applications, as well as having the capability for expansion and scalability. Using GA with

Python is now simpler than ever, providing a work-alike replacement for Python's NumPy module, the de facto standard for numerical computing in Python. The new Python module, Global Arrays in NumPy (GAiN), allows the development and debugging of serial NumPy codes which can later scale on more capable clusters or supercomputers - often by changing only one line of Python code.

The GA tutorial and related paper presentation at this year's Python for Scientific Computing Conference (SciPy) generated enough buzz that SciPy's institutional sponsor, Enthought Inc., is considering incorporating GA and GAiN into its Python distribution. PNNL is pursuing funding opportunities and looking to expand our user base.

More information: Daily JA, and RR Lewis. 2011. "Using the Global Arrays Toolkit to Reimplement NumPy for Distributed Computation." In *SciPy* 2011. PNNL-SA-80943, Pacific Northwest National Laboratory, Richland, WA. http://www.archive … mentNumpyFor

Daily JA, et al.. 2011. "High Performance Computing in Python using NumPy and the Global Arrays Toolkit." Presented by Jeff Daily at *SciPy* 2011, Austin, TX on July 12, 2011. PNNL-SA-81329. http://conference. … als.php#jeff

Daily JA, et al. 2011. "Overview of the Global Arrays Parallel Software Development Toolkit." PyCon 2011.

Daily JA, et al. 2011. "PyCon 2011 Lightning Talk: The Global Arrays Parallel Programming Toolkit." PyCon 2011.

Provided by Pacific Northwest National Laboratory

"Python bindings snake into global arrays toolkit." September 26th, 2011. http://www.physorg.com/news/2011-09-python-snake-global-arrays-toolkit.html