Welcome to the Python in HPC Webinar

June 7, 2017
1:00 pm EDT

You may ask your questions directly in the WebEx application using the Chat function or if you cannot access the chat capability inside of WebEx, you may type your questions in this Google doc.

After the webinar, please take our survey: http://bit.ly/2qY3Yj5 Thanks!

Can cProfile profile python programs that use MPI?
- Yes, it can generally produce profiling output per rank.

Thanks. Is there a specific API for producing the output per rank, or is it simply
`mpirun -np N python -m cProfile -o output.prof script.py`?
- That input string is generally sufficient, it should produce ranked output by default.

Great. Thanks!

Is it planned that OLCF and ALCF support Jupyter notebook, like NERSC does at https://jupyter-dev.nersc.gov/ ?
- Both the OLCF and ALCF, have an interest in providing Jupyter instances to users. Currently it is possible to run Jupyter in a *limited* sense on Rhea (contact belhornmp@ornl.gov for details if interested). Using jupyter on the OLCF Crays Titan and Eos is currently unsupported and will likely remain so for the foreseeable future. The OLCF is building an open enclave in which the NCCS will be better able to deploy services similar to Jupyterhub that do not currently mesh with the OLCF's moderate security mandate.

OK, thanks for the answer!
- You're welcome. If you have further questions about Jupyter on Rhea, feel free to send me an email at `belhornmp@ornl.gov`

I am now using joblib for parallel works. What are the major disadvantages of joblib compared to mpi4py?
- HPC still tends to rely on MPI for interprocess communication in such a way that the greatest utilization of dedicated HPC resources requires the use of MPI. In particular, many HPC resources have optimizations at the hardware and firmware level for using

MPI over specific interconnects or otherwise suppress internode communication over TCP/UDP ethernet channels. That said, many software of interest to scientific computing, such as popular machine learning frameworks, are moving beyond MPI.

Is it planned that NERSC/OLCF/ALCF support Numba (http://numba.pydata.org/) as an alternative to Cython/f2py ?
- Answered in conversation. Transcript needed. Otherwise see the webinar recording.

Questions asked in WebEX Chat Session:

from SR to Everyone:
- What are bindings? C extensions, can you give a little more background?
from WS to Everyone:
- Bindings are interfaces for libraries and code written in another langauge. Extensions are another way of saying modules or libraries.

from TJD to Everyone:
- One library that wasn't mentioned is numexpr which may be of interest
from Matt Belhorn (OLCF) to Everyone:
- @TJD, good suggestion

from LL to Everyone:
- Do you think vtune or allinea is better?
from Matt Belhorn (OLCF) to Everyone:
- @LL, I prefer allinea in general but vtune has some targeted support for python
from B to Everyone:
- I'm from Allinea. We do not yet support python but support will be coming soon.
from WS to Everyone:
- @B I had decent luck with CPython under MAP/DDT, though things are rather mangled making finding function names a pain.

from TJD to Everyone:
- Are there any plans for supporting a JupyterHub type interface to these HPC systems?
from Matt Belhorn (OLCF) to Everyone:
- @TJD, several sites are working towards supporting Jupyter on major compute resources, though there are some challenges.
from D to Everyone:

- We do have an experimental Jupyter interface at NERSC: http://www.nersc.gov/users/data-analytics/data-analytics/web-applications-for-data-analytics/

from TJD to Everyone:
- @D Thank you for the resource

from SE to Everyone:
- What are your thoughts on using PyCharm over Jupyter?

\# Answered in the voice recording by the speakers.

from WI to Everyone:
- Does mpi4py support Pandas dataframes?

from Matt Belhorn (OLCF) to Everyone:
- @WI Pandas dataframes can generally be pickled and would be compatible with mpi4py

from SR to Matt Belhorn (OLCF) (privately):

from KS to Sherry Ray (privately):
- what is pickling?

from Matt Belhorn (OLCF) to Everyone:
- Pickling, for those who are not sure, refers to serializing an instance of a python object as it exists in memory. The object is translated into a format such as JavaScript Object Notation - JSON that preserves the rich structure of the class instance. The serialized form can then be saved to disk, transmitted, etc and restored into memory as a clone of the original object at a later time even in a different python runtime instance through the process of `unpickling`. See python.cPickle, python.pickle, or the primitive python.marshal for details about pickled python objects.

from DR to Everyone:
- https://drtiresome.com/2016/08/23/build-and-install-mpi-parallel-hdf5-and-h5py-from-source-on-linux/

from DR to Everyone:
- Possibly helpful

from Matt Belhorn (OLCF) to Everyone:
- @DR, the backup slides also contain some information on building mpi4py and h5py on cross-compile environments like many Crays

from DR to Everyone:
- Awesome, thanks!

from MO to Everyone:
- can you say a word about using containers (e.g. Docker) when writing Python applications (for re-use

from Matt Belhorn (OLCF) to Everyone:

- @MO, docker is generally great for portability, but not many HPC resources have extensive support for it outside of shifter and singularity

from MO to Everyone:
- ok, thanks!