Simple Task Parallelism with Python

Sheri Mickelson
Software Engineer, NCAR

Introduction to PyNIO and Related Python Tools for Geoscientific Data Analysis
AMS 96th Annual Meeting

Sunday, January 10, 2016
CISL/TDD/ASAP Group

• Our group focuses on
  – Improving the scalability and performance of NCAR applications
  – Efficient use of accelerator technology
  – Workflow and I/O
ASAPTools

• Contains wrappers around mpi4py functionality

• Requires:
  – MPI (any flavor)
    • This needs to be installed and in your path before proceeding with the Python installs
  – mpi4py

• Can be installed with
  – pip install [--user] ASAPTools
Parallelism

What can be executed at the same time?
For Example

Calling the same NCL script for different input files ...

ncl 'srcFileName=“file_1.nc”' regrid.ncl
ncl 'srcFileName=“file_2.nc”' regrid.ncl
ncl 'srcFileName=“file_3.nc”' regrid.ncl
ncl 'srcFileName=“file_4.nc”' regrid.ncl
For Example

If run serially, the cost would be 4 \times N
If run in parallel and each NCL call had its own MPI task, the cost would be N
Another Example

Calling different independent NCL scripts ...

```
ncl plot_SALT.ncl
ncl plot_ADVS.ncl
ncl plot_UES.ncl
ncl plot_UVEL.ncl
ncl plot_VVEL.ncl
ncl plot_BSF.ncl
ncl plot_TAUX.ncl
ncl plot_TAUY.ncl
```
Another Example

Calling different independent NCL scripts ...

```
ncl plot_SALT.ncl          →  MPI Task
ncl plot_ADVS.ncl          →  MPI Task
ncl plot_UES.ncl           →  MPI Task
ncl plot_UVEL.ncl          →  MPI Task
ncl plot_VVEL.ncl          →  MPI Task
ncl plot_BSF.ncl           →  MPI Task
ncl plot_TAUX.ncl          →  MPI Task
ncl plot_TAUY.ncl          →  MPI Task
```

All of these lines can all be executed at the same time because they are all independent of each other.
Steps to run a Python script in parallel with ASAPtools

• Add to the script:
  1. from asaptools import simplecomm, partition
  2. comm = simplecomm.create_comm(serial=False)
  3. local_list =
     comm.partition(global_list, func=partition.EqualStride(), involved=True)

• Execute with ‘mpirun –n <# of tasks> python script.py’
DEMO: EXAMPLE ON HOW TO PLOT ALL OF THE 3D VARIABLES IN A FILE IN SERIAL AND IN PARALLEL
Other operations within ASAPTools

• Other MPI Support
  – Multiple partitioning methods
    • EqualLength, EqualStride, SortedStride, and WeightBalanced
  – Sends/Recv
  – Split Communicators
  – Rank/Master/Size Query

• Timers
Questions?

• ASAPTools
  – pip install [--user] ASAPTools

• mpi4py
  – pip install [--user] mpi4py
  – https://bitbucket.org/mpi4py/mpi4py

• MPI
  – Building from source instructions: https://mpi4py.scipy.org/docs/usrman/ap...building-mpi-from-sources
  – MPICH: http://www.mpich.org/
  – OpenMPI: https://www.open-mpi.org/