Scripting for HPC

Drew Dolgert
Cornell Center for Advanced Computing
December 8, 2010
What’s a scripting language?

• Python
• Perl
• R
• Lua
• Tcl
• Ruby
• Bash, Csh

What scripting language is more popular than all of these?
What’s It Look Like?

• Go to ~/labs/python
• Look at simple.py.
• Run it with python simple.py 100

What do you notice about how it differs from C, Fortran?
Dynamic = Built to Play

- Graphics – Matplotlib, VTK, gnuplot
- Numerics – Numpy, Scipy
- Data Transformation – XML, binary, HDF
- Networking – MPI, TCP/IP, web services

Can’t Fortran do these things?

12/9/2010
Luna the rabbit says RELAX.
Exercise: Pure Math Speed

- $ cd python
- Add python and numpy modules.
- $ time python pure_invert.py 100
- $ time python numpy_invert.py 100
- Is the code reasonable?
YOU ARE BEAUTIFUL
Top Level Logic

Integration, Larger loops, Middle Stuff

Innermost Subroutines
Replacing Top Level

Initialize()
Decide what you store.
Loop until done:
  Integrate()
Write final results()
Write summary results()
File with Hints for Wrapping

Wrap Program

C headers

Scripting Code

Glue Library

Compiled Library
from mpi4py.mpi_c cimport MPI_Comm, MPI_COMM_WORLD, MPI_Comm_split, MPI_Comm_rank

cdef extern from "library.h":
    void lammmps_open(int argc, char** argv,
                        MPI_Comm comm, void** lammmps_ptr)
    void lammmps_close(void* lammmps_ptr)
    void lammmps_file(void* lammmps_ptr, char* file)
    char* lammmps_command(void* lammmps_ptr,
                           char* command)
Almost All Fortran

Parameter File

Initialization Subroutine

Occasional Visualization
C or Fortran

Glue

R, Python, Lua
Compiled keeps track of state of interpreter.
Try Exercise with Lua
Let’s do the whole thing in R, in parallel, on Ranger.

Write almost all of it, even integration, in Python.
Beazley, 1997

- You can write inner loops in Python if you…

Always manipulate sets of atoms.
# A function written in Python
from SPaSM import *
def run(nsteps, Dt, freq):
    for i in xrange(0, nsteps):
        integrate_adv_coord(Dt)
        boundary_periodic()
        redistribute()
        force_eam()
        integrate_adv_velocity(Dt)
        if (i % freq) == 0):
            output_particles('Dat'+str(i))
Two Ways for Pure R

- Leave R untouched, but run it thousands of times.
- Use Rmpi library to make parallel R.

Not fast, but scalable, malleable.
R Labs

• Run Serial R Many Times
• Run an R MPI job

Both easier as a group.
Gratefully Borrowed

http://www.flickr.com/photos/jvuokko/374781736/

http://www.flickr.com/photos/lovenotfear/490469619/