High Performance Computing Modernization Program (HPCMP) Summer 2011 Puerto Rico Workshop on Intermediate Parallel Programming & Cluster Computing

in conjunction with the National Computational Science Institute (NCSI)/ SC11 Conference

SHODOR

EARLHAM COLLEGE

Jointly hosted at Polytechnic U of Puerto Rico and U Oklahoma

and available live via videoconferencing (streaming video recordings coming soon)

HAMPTON UNIVERSITY

Widener University

Sponsored by DOD HPCMP, SC11/ACM, NCSI and OK EPSCoR

Intermediate Parallel Programming & Cluster Computing

I/O Libraries - netCDF

Sponsored by
DOD HPCMP,
SC11/ACM,
NCSI and
OK EPSCoR

Josh Alexander, University of Oklahoma
Ivan Babic, Earlham College
Ken Gamradt, South Dakota State University
Andrew Fitz Gibbon, Amazon.com
Mobeen Ludin, Earlham College
Tom Murphy, Contra Costa College
Henry Neeman, University of Oklahoma
Charlie Peck, Earlham College
Stephen Providence, Hampton University
Jeff Rufinus, Widener University
Luis Vicente, Polytechnic University of Puerto Rico
Aaron Weeden, Earlham College

Sunday July 31 – Saturday August 6 2011

This is an experiment!

It's the nature of these kinds of videoconferences that **FAILURES ARE GUARANTEED TO HAPPEN! NO PROMISES!**

So, please bear with us. Hopefully everything will work out well enough.

If you lose your connection, you can retry the same kind of connection, or try connecting another way.

Remember, if all else fails, you always have the toll free phone bridge to fall back on.

H.323 (Polycom etc)

If you want to use H.323 videoconferencing – for example, Polycom – then:

- If you ARE already registered with the OneNet gatekeeper, dial 2500409.
- If you AREN'T registered with the OneNet gatekeeper (which is probably the case), then:
  - Dial **`164.58.250.47`**
  - When asked for the conference ID, enter:

    `#0409#`

Many thanks to Roger Holder and OneNet for providing this.

H.323 from Internet Explorer

From a Windows PC running Internet Explorer:

1. You **MUST** have the ability to install software on the PC (or have someone install it for you).
2. Download and install the latest Java Runtime Environment (JRE) from here (click on the Java Download icon, because that install package includes both the JRE and other components).
3. Download and install this video decoder.
4. Start Internet Explorer.
5. Copy-and-paste this URL into your IE window:
   http://164.58.250.47/
6. When that webpage loads, in the upper left, click on "Streaming".
7. In the textbox labeled Sign-in Name, type your name.
8. In the textbox labeled Conference ID, type this:
   0409
9. Click on "Stream this conference".
10. When that webpage loads, you may see, at the very top, a bar offering you options. If so, click on it and choose "Install this add-on."

EVO

There's a quick description of how to use EVO on the workshop logistics webpage.

Phone Bridge

If all else fails, you can call into our toll free phone bridge:

1-800-832-0736

* 623 2874 #

Please mute yourself and use the phone to listen.

Don’t worry, we’ll call out slide numbers as we go.

Please use the phone bridge **ONLY** if you cannot connect any other way: the phone bridge is charged per connection per minute, so our preference is to minimize the number of connections.

Many thanks to OU Information Technology for providing the toll free phone bridge.

Please Mute Yourself

No matter how you connect, please mute yourself, so that we cannot hear you.

At ISU and UW, we will turn off the sound on all conferencing technologies.

That way, we won't have problems with echo cancellation.

Of course, that means we cannot hear questions.

So for questions, you'll need to send some kind of text.

Questions via Text: Piazzza

Ask questions via:

http://www.piazza.com/

All questions will be read out loud and then answered out loud.

NOTE: Because of image-and-likeness rules, people attending remotely offsite via videoconferencing **CANNOT** ask questions via voice.

Thanks for helping and sponsoring!

- OSCER operations staff (Brandon George, Dave Akin, Brett Zimmerman, Josh Alexander, Patrick Calhoun)
- Debi Gentis, OU
- Kevin Blake, OU IT (videographer)
- OU School of Electrical & Computer Engineering (LittleFe buildout)
- James Deaton and Roger Holder, OneNet
- Luis Vicente and Alfredo Cruz, Polytechnic U of Puerto Rico
- Omar Padron, Kean U
- Scott Lathrop, SC11 General Chair
- Donna Cappo, ACM
- Bob Panoff, Jack Parkin, Joyce South, Shodor Education Foundation Inc
- Jerry Malayer and Jim Wicksted, Oklahoma EPSCoR
- Dept of Defense High Performance Computing Modernization Program

This is an experiment!

It's the nature of these kinds of videoconferences that **FAILURES ARE GUARANTEED TO HAPPEN! NO PROMISES!**

So, please bear with us. Hopefully everything will work out well enough.

If you lose your connection, you can retry the same kind of connection, or try connecting another way.

Remember, if all else fails, you always have the toll free phone bridge to fall back on.

Outline

- I/O Challenges
- Number and Text Representations
- Portable Binary Output
- NetCDF

I/O LIBRARIES

netCDF

Do Not Reinvent the Wheel

I/O Challenges

- **I/O plays a very important role in Scientific Computing**

  Ideas -> Computation -> Output (Data)

- **However,**

  Data != Information

I/O Challenges

- **Example:**

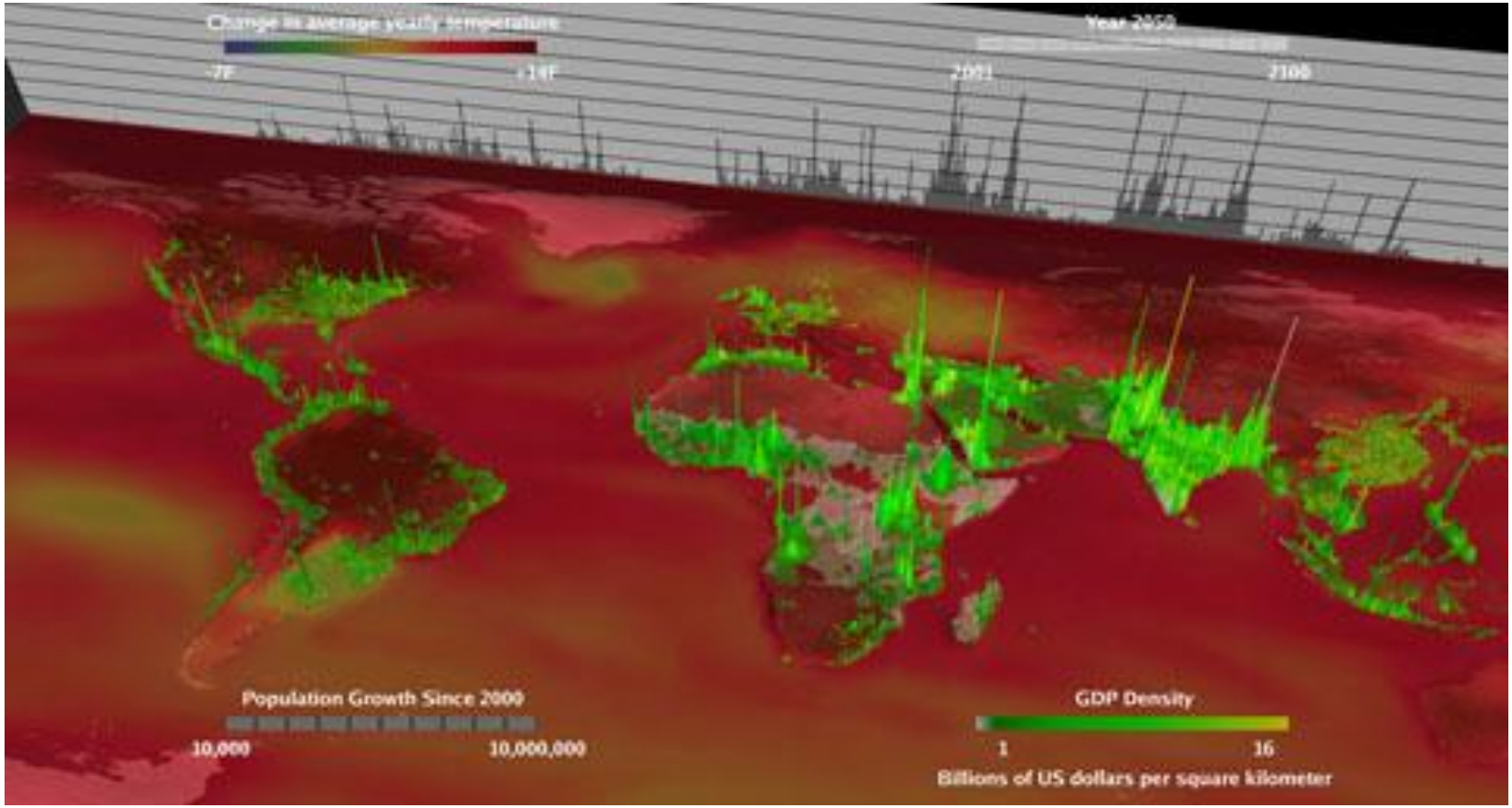
A set of data

11.0, 3.24, 4.5, 8.8, 7.88, 74.2, 87.3, 12.3, 8.77,
2.4, 3.25, 9.1, 82.3, 93.2, 87.2, 24.5, 33.1, 87.2,
4.4, 6.11, 22.1, 9.8, 1.04, 6.73, 11.43, 76.3, 9.61,
etc. etc. etc.

??

I/O Challenges

- How about (Ref [2])



I/O Challenges

- Thus,

Data must be analyzed/interpreted/visualized in order to be "meaningful" (information)

I/O Challenges

- However,

I/O also creates challenges in Scientific Computing

  - I/O is expensive - takes much more time than computation
  - I/O is represented differently in different computers

Text Representation

PRINTING OUTPUT

- Text (ASCII) : 7 bits
- Extended ASCII : 8 bits = 1 Byte (each character = 1 Byte)

- Examples:
  - ABC => 3 Bytes
  - 08/03/2011 => 10 Bytes
  - 3.14159265358979323846 => 20 Bytes
  - 6.0221415e+23 => 13 Bytes

ASCII = American Standard Code for Information Interchange

Machine Interpretation

C/C++ :

```
int x = 8;
float y = 12.0;
double z = 20.0;
```

Fortran:

```
integer x = 8
real y = 12.0
double precision z = 20.0
```

(assume 32 bits computers)

x is 4 bytes

y is 4 bytes

z is 8 bytes

Binary versus Text

- 8 Bytes (Binary) is less than 13 Bytes, 20 Bytes, etc (Text).
- Output data can be written as binary (internal representation – opaque from outside) in C/C++, Fortran, etc.
- Save time
- Save storage

Issues with Binary Output

- Readability issue: We cannot read binary
  - No problem, we could always change the format binary <-> text

- Portability Issue: Different computers represent numbers in different ways
  - Data might not be used by different computers
  - Data might become obsolete over the years

Portable Binary Output

- NCSA-HDF (National Center for Supercomputing Applications - Hierarchical Data Format)
  - http://www.hdfgroup.org
- Unidata – netCDF (Network Common Data Form)
  - http://www.unidata/ucar.edu/software/netcdf
  - Atmospheric science/climate modeling

netCDF

- Portable binary I/O
- Free, open source
- Run on Linux, Windows, MacOS
- Interface with C/C++, Fortran, Java, Matlab, Perl, Python, etc.
- Multidimensional Array-Oriented Data Access

netCDF

- Available software to access, manipulate, and visualize netCDF data:
  - Free: OpenDX, Ferret (computer visualization for oceanographers), etc.
  - Commercial: Mathematica, Matlab, etc.
- Contains Metadata (i.e. data about data – file contains title, units, source, dimensions, type, etc.)
- The whole or a subset of the data can be access efficiently
- Parallel I/O is available

netCDF

- Say we have a model to compute Temperature (T) and Pressure (P) at a specific point (Latitude and Longitude)
- Idea -> Computation -> Output
- In the output we would like to put
  - Units of T (degrees Celsius), P (hPa), Lat (degrees), Lon (degrees)
  - Data of Latitude, Longitude, T(Lat, Lon), P(Lat, Lon)

Note: Latitude runs North-South, Longitude runs East-West

netCDF

- netCDF file name always ends with “nc” :

  example.nc

- ncdump creates TEXT file (all or partial)

  ncdump example.nc > example.txt

- ncgen creates netCDF file (e.g. to be edited, modified)
  - Edit example.txt (using vi, nano, emacs, etc.)
  - **ncgen –o example.nc example.txt**

Example of netCDF file

```
netcdf sfc_pres_temp {
dimensions:
        latitude = 6 ;
        longitude = 12 ;
variables:
        float latitude(latitude) ;
                latitude:units = "degrees_north" ;
        float longitude(longitude) ;
                longitude:units = "degrees_east" ;
        float pressure(latitude, longitude) ;
                pressure:units = "hPa" ;
        float temperature(latitude, longitude) ;
                temperature:units = "celsius" ;
```

Example of netCDF file

```
data:
 latitude = 25, 30, 35, 40, 45, 50 ;
 longitude = -125, -120, -115, -110, -105, -100, -95, -90, -85, -80, -75, -70 ;

 pressure =
  900, 906, 912, 918, 924, 930, 936, 942, 948, 954, 960, 966,
  901, 907, 913, 919, 925, 931, 937, 943, 949, 955, 961, 967,
  902, 908, 914, 920, 926, 932, 938, 944, 950, 956, 962, 968,
  903, 909, 915, 921, 927, 933, 939, 945, 951, 957, 963, 969,
  904, 910, 916, 922, 928, 934, 940, 946, 952, 958, 964, 970,
  905, 911, 917, 923, 929, 935, 941, 947, 953, 959, 965, 971 ;

 temperature =
  9, 10.5, 12, 13.5, 15, 16.5, 18, 19.5, 21, 22.5, 24, 25.5,
  9.25, 10.75, 12.25, 13.75, 15.25, 16.75, 18.25, 19.75, 21.25, 22.75, 24.25,
    25.75,
  9.5, 11, 12.5, 14, 15.5, 17, 18.5, 20, 21.5, 23, 24.5, 26,
  9.75, 11.25, 12.75, 14.25, 15.75, 17.25, 18.75, 20.25, 21.75, 23.25, 24.75,
    26.25,
  10, 11.5, 13, 14.5, 16, 17.5, 19, 20.5, 22, 23.5, 25, 26.5,
  10.25, 11.75, 13.25, 14.75, 16.25, 17.75, 19.25, 20.75, 22.25, 23.75,
    25.25, 26.75 ;
```

Thanks for your attention!

Questions?

References

[1] Neeman, H. “Supercomputing in Plain English” www.oscer.ou.edu/education.php
[2] http://dart.ncsa.uiuc.edu/avl/AtmosphericSciences.html
[3] http://www.unidata.ucar.edu/software/netcdf/