A Training Roadmap for New HPC Users

Mark Richards, Scott Lathrop
mdrichar@ncsa.illinois.edu
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Outline

• Motivation for the Roadmap
• Orienting New Users to the World of HPC
• Survey of Resources
• Keys for Success Going Forward
Most HPC users are formally trained as scientists rather than programmers

Are the majority of your current programming language skills self-taught?

As HPC becomes more pervasive, the community will become more diverse
A common trait of HPC users: They do not know what they do not know

“There are known knowns; there are things we know that we know. There are known unknowns; that is to say there are things that we now know we don’t know. But there are also unknown unknowns – there are things we do not know we don’t know.”

--Donald Rumsfeld
A training roadmap is more than a list of resources or FAQ

- How can I add or remove a user from an existing TeraGrid allocation?
  Only principal investigators, co-principal investigators, or allocation managers can add or remove users from a TeraGrid allocation.

- How do I apply for a new TeraGrid allocation?
  Note: A TeraGrid allocation is not an account. An account is the specific resource that you are using.

- What’s the quickest way for new users to access the TeraGrid?
  Note: The information below assumes you have finished all the initial steps.

- On the TeraGrid, what is a Science Gateway?
  A Science Gateway is a community-developed set of tools, applications, and data available to the research community.

- How do I get a certificate proxy?
  On this page: Introduction to obtaining a grid certificate proxy: IU, NCSA, and TeraGrid.

- On the TeraGrid, how much long-term storage can I get?
  For information about storage quotas and purge policies at each TeraGrid site.

- On the TeraGrid, how can I get my data transfers to run faster?
  While many of the factors that affect data transfer rates are intrinsic to the network, there are some steps you can take to optimize transfer performance.

- On the TeraGrid, what is a Roaming allocation?
  Note: Roaming allocations on the TeraGrid are no longer available for new users.

- On the TeraGrid, what types of allocations are available?
  Note: A TeraGrid allocation is not an account. An account is the specific resource that you are using.

- On the TeraGrid, how are compute jobs charged?
  Note: Indiana University will retire Big Red from the TeraGrid July 31, 2011.
Our roadmap is a web-based flowchart of key concepts in the HPC space.
Each node in the flow chart links to a brief description and a list of resources

- Memory and Storage Hierarchy
  - SCIPE: Tyranny of Storage Hierarchy
- Parallel Overhead
  - Communication/Computation Ratio
  - Tradeoffs: Communication vs. Redundant Computation
- High Throughput Computing
  - SCIPE: High Throughput Computing
- Numerical Analysis Issues
- Networking Basics
  - Latency
    - Latency is the time that elapses between the request of message and when the (first part of) the message is received.
- Getting Started Guides
  - CL-Tutor
  - Ohio Supercomputing Center: HPC Resources
  - NCSA Getting Started Guide
  - NICS: Quick Start
  - NICS: New to HPC

View article
Initial topics have been selected in consultation with the HPC community.
Resources include tutorials, slide sets, glossary, articles, books, and videos

Tutorials

Books & Articles

Videos

Glossary

- All Gather
- All to All
- Bandwidth
- Blocking Send
- Broadcast
- Cache
- Latency
- MPI
- Node
- PerfSuite
- Shared Memory
- X10
A key goal is to convey concise explanations of key concepts early

“Education is the process of telling smaller and smaller lies.”

quoted by J.R. Deller, Jr.
Example: The basics of cache can be explained in a minute or an hour.
We need to capture the attention of new HPC users while they are still new

TeraGrid Access Information

Welcome to TeraGrid!
As part of receiving your first project allocation on TeraGrid (as PI or co-PI) or being added as a user to your first project allocation on TeraGrid, you have been given the following TeraGrid-wide username and password. This is your TeraGrid username for all future allocations and projects; you will not need to wait for postal mail in the future. Please keep your user information up to date in the TeraGrid User Portal.

Your TeraGrid Login
Username: mrichard Password:

Treat passwords as secure information.
Do not share your password with anyone, including TeraGrid Staff.
Do not send passwords via e-mail.

Your User Information
Mark Richards
University of Illinois, Urbana
NCSA
4103 NCSA
1205 W Clark St
Urbana, IL 61801
United States

Please see http://hpcuniversity.org to orient yourself to the world of high performance computing, to learn about common problems and how you can avoid them, and to find helpful training resources

TeraGrid User Portal — https://portal.teragrid.org/
TeraGrid Access Documentation — https://www.teragrid.org/web/user-support/access
We need the community to suggest topics and resources for inclusion

Use this form to submit new resources or metadata revisions for the HPCU catalog.

HPCU Resource Documentation Form

Title: *

Url: * The URL where your project/resource is stored, i.e. http://www.shodor.org

Description: * A free-text account of a resource. Used in displaying search results

Creator Information:

Creator: (one entry per line)
We need HPC users to provide feedback on the quality of resources

Visualization Resources

- CITutor Course: Introduction to Visualization
- Ranger Virtual Workshop: Large-scale Visualization
- Ranger Virtual Workshop: Paraview

**Paraview** is a visualization application highly capable for computational fluid dynamics and other subjects. It is open source and can run in parallel on Ranger. This module includes a lab which covers visualization of a sample dataset both on a local computer and on TACC resources.

- Read 14 User Reviews
- Gallery
  - Ranger: VisIt
- Supercomputing in Plain English: Math Libs, I/O, Visualization
- Difference between VisIt and ParaView

Write a Review
Summary

• The Training Roadmap helps new users—particularly science and engineering graduate students—navigate the space of high performance computing

• Concise explanations of fundamental concepts help users quickly understand what it is they do not yet know

• Community feedback will help to identify the best resources, fill gaps, and adapt to changes

http://hpcuniversity.org