A CURRICULUM FOR PETASCALE COMPUTING

Thomas Hacker, Gary Bertoline

Teragrid ‘08 Conference, Las Vegas, NV

Outline

- Introduction and Motivation
- Petascale Computing Challenges
- Curriculum for Petascale Computing
- Desired Outcomes
- Conclusion

Introduction and Motivation

- Petascale computing represents a new frontier in supercomputing
- Significant challenges to overcome in several areas
  - Engineering
  - Technology
  - Computer science
  - Visualization
  - Software engineering

The Petascale Challenge

- Some examples to illustrate the magnitude of the challenge inherent in Petascale Computing
- Data Movement
- Power Consumption

Data Movement Challenge

- CPUs must be kept fed with Data/Instructions
  - If the storage system cannot keep up with the CPU, it will stall, wasting power, time, and money
- Using a conservative metric of 200 MB/sec of I/O bandwidth per TF of computation
  - 1 PF system will require 195 GB/sec of I/O
  - Network bandwidth would be 1.562 Terabits/sec
- How will we build and operate data systems with these capabilities?

Power Consumption

- Frequency scaling processors can consume less power at slower clock speeds
- In a petascale system comprised of 10,000 processors, scaling back even 5% of the cores to half speed would save a lot of power
- How do we devise codes and operating systems support to reduce power consumption?

The Problem

- Unmet need for practitioners skilled in several areas:
- Deploying and building petascale computing and storage systems
- Operating and managing petascale systems
  - Systems administration, tuning and configuration
- Effectively utilizing petascale systems
  - Skilled developers
  - Domain specialists who can apply HPC technologies to solving difficult problems

Petascale Curriculum

- To meet the demands of petascale computing, we propose that the high performance computing community needs an expanded curriculum that expands training and education *beyond* current efforts in several key areas.
  - Parallel computation and algorithm development
  - Petascale systems architecture and operating systems
  - Petascale systems operations and systems administration
  - Petascale visualization
  - Science gateway development

Current HPC Curriculum Efforts

- Several trends and forces are driving current efforts to develop CI and HPC curriculum:
  - Industry demand for skilled practitioners
    - Potential impact on HPC technologies on U.S. competitiveness
  - Pervasive availability of key components
  - Opportunity to help shape an emerging field

Industry Demand

- Growing use of high performance computing and supercomputing for research and development in engineering and science
- Automobile Industry
  - Full vehicle design, optimization and testing
  - General Motors estimated 40% reduction in engineering costs and design cycle time reduced by 60%
- Boosting oil and gas recovery from fields
  - Reanalyzing archived seismic data with new algorithms
- Potential to enhance the competitiveness of American industries in a global marketplace

Source: Council on Competitiveness

Pervasive Availability of Key Components

- Trend #1: Rising power and decreasing cost of commodity hardware
  - Powerful, low cost components available for HPC designer
  - Commodity based clusters now dominate the Top500 list of the fastest supercomputers in the world
- Trend #2: Open source software
  - Free or low-cost, high quality operating systems, compilers, and applications
  - Linux, MPI, BLAST, ….
  - Can be used to solve difficult science and engineering problems
- Combination of these trends can potentially help manufacturers, engineering, architects, and designers to deliver much better products and services at a lower cost

Curriculum Today at Purdue

- Partnerships with industrial colleagues in development
  - Bioinformatics and Discovery Systems online Spring 2008
  - HPC Lab coming online in August 2008
- New courses
  - CIT 499M High Performance Computing Systems
  - CIT 581D Research Computing and Data Analysis
  - CIT 581C Introduction to Computational Life Sciences
  - CIT 581B BioMed Systems Analysis and Design
  - CIT 499x/581x Health and Biomedical Informatics
- New Area of Specialization
  - MS in Scientific & Information Visualization
- In development
  - CIT 581M Advanced High Performance Computing Systems
  - CIT 499 Parallel Data Systems

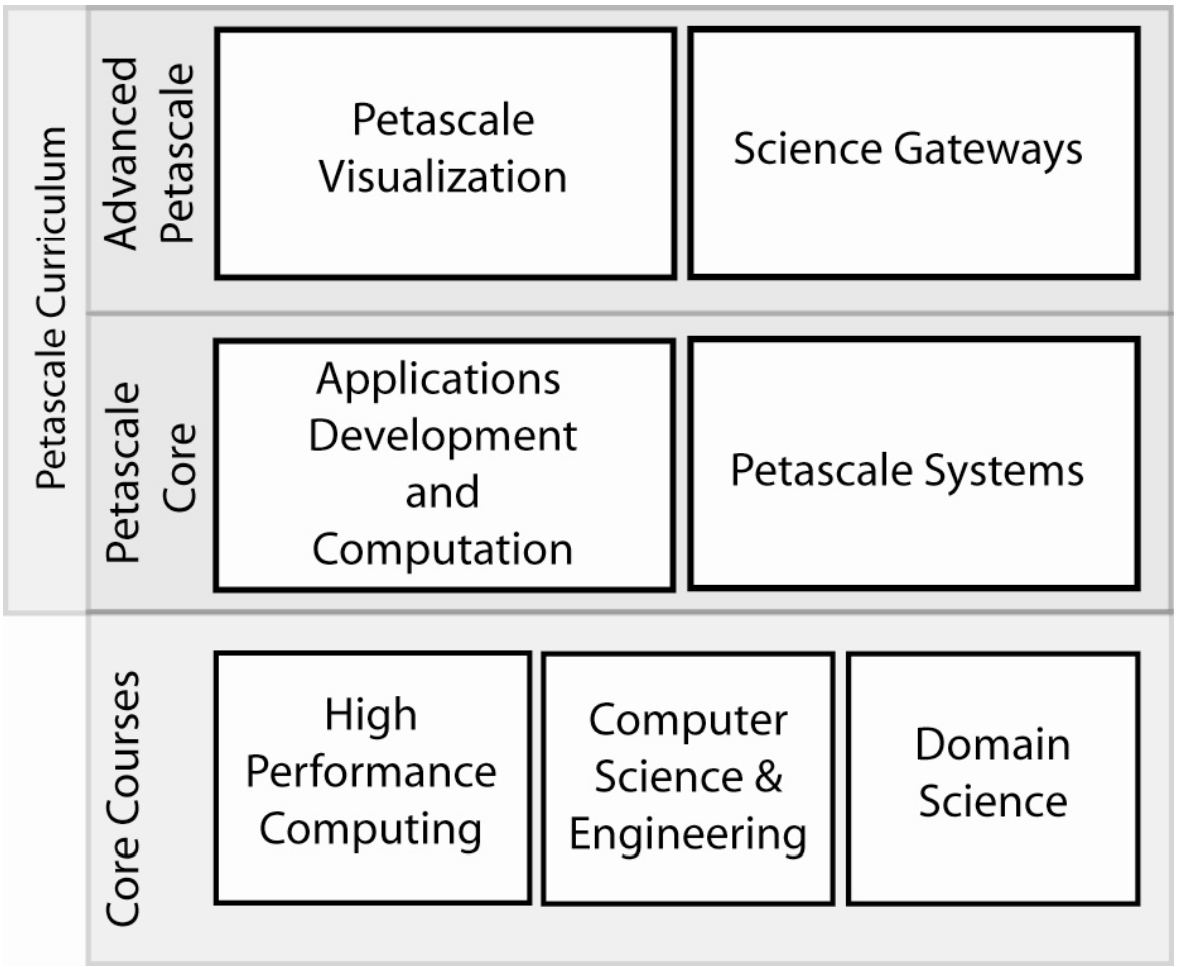
Petascale Curriculum

- Developing a petascale curriculum
  - Build on existing curriculum efforts
  - Develop partnerships with petascale centers
  - Example: Great Lakes Petascale Consortium
    - Focused on developing and using upcoming NSF funded Petascale system
    - Led by NCSA

Petascale Curriculum Goals

- Provide training to address the immediate problems of petascale computing
- Provide course content along with project experience that will remain relevant over a long period of time and establish a foundation for future learning and research
- Build on existing practice and knowledge in high performance computing

Petascale Curriculum Map



Proposed Petascale Curriculum

- Core courses
  - Training in science, technology, engineering, mathematics, or computer science
  - Additional core courses needed for further progression in any of the three petascale curriculum areas. This set of core courses includes: grid computing, numerical analysis and statistics, algorithms, high performance computing systems (focusing on cluster architectures), and database systems.
- Specialization Tracks
  - Applications Development and Computation
  - Petascale Systems
  - Petascale Visualization
  - Science Gateways

Applications Development and Computation

- A course sequence focused on content necessary for application development and petascale computation
  - Numerical analysis, serial and parallel algorithms, operating systems, networking and communications, software engineering, and training in biology, physics, chemistry, or one of the engineering disciplines will be needed to prepare students to successfully develop efficient high performance applications for petascale systems.
  - Special emphasis on Multicore processors, reliability and fault tolerance,

Petascale Systems

- Petascale Systems
  - The design, operation, and administration of petascale systems.
- This area has five main components
  - Data center design and operation
  - Operating systems
  - Systems design
  - Communications
  - Data storage systems

Petascale Visualization

- Petascale Visualization
  - Training in parallel visualization and rendering techniques
  - GPU processor accelerated visualization techniques
  - In-situ data processing and visualization
  - Remote, collaborative data analysis and visualization

Science Gateways

- Science Gateways
  - Computer graphics; human interface design; web programming; educational technology and integration with learning management systems; and scientific workflows and database systems
  - Focus on usability as well as maximizing the education and research impact of science gateways for communities ranging from middle school science courses to global science collaborations involving thousands of researchers using cyberinfrastructure.

Desired Outcomes

- Develop a cohort of students equipped to develop and exploit the capabilities of petascale systems
- Build the necessary groundwork for achieving the next frontier of HPC
  - Exascale computing
  - Recall that 1 TF was achieved only a decade ago
  - It will take a few years to develop the base of courses and expertise needed to reach for Exascale systems
- Develop new capabilities to address pressing issues in Petascale computing
  - Power efficient applications
  - Reliability and usability
  - Data management

Conclusions

- Petascale computing is a new frontier in high performance computing that poses significant challenges to design effective and efficient systems, applications, and operations
- To meet this challenge, we must develop an enhanced curriculum that builds on existing courses to address petascale computing needs
- Through the development of this curriculum, we will help develop a cohort of students with the skills and training to effective use petascale systems, and set the stage for the future development of Exascale systems.