GridForce:

A Multi-tier Approach to Prepare our Workforce for Grid Technology

Bina Ramamurthy
CSE Department
University at Buffalo (SUNY)
201 Bell Hall, Buffalo, NY 14260
716-645-3180 (108)
bina@cse.buffalo.edu
http://www.cse.buffalo.edu/gridforce
Partially Supported by NSF DUE CCLI A&I Grant 0311473

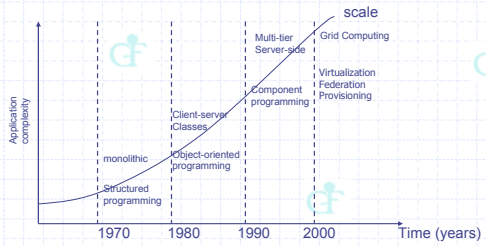
Introduction

- We present an adaptation of the upcoming grid technology in CS-based curriculum.
- Courses span multiple tiers:
  - CS undergraduate senior level (CSE486, CSE487)
  - CS graduate entry level (CSE586, CSE587)
  - Entry level scientists and engineers
  - Seminars to industry
- Goal is to improve technical preparedness of our workforce for grid technology.

Topics for Discussion

- What is grid technology? (General, Technical)
- Why grid technology?
- Adaptation of Grid Technology to CS Curriculum
- GridForce Project
- Courses: Curriculum CSE4/586, CSE4/587
- Lab Exercises: problem, approaches to solution, code base for solution
- Fundamental concepts covered
- Technologies and tools covered
- Preliminary Assessment of Effectiveness of Adaptation
- Grid infrastructure
  - Reusing old hardware (SparcGrid)
  - Grid with newer hardware (LinuxGrid)
- Industrial outreach
- Challenges in Adaptation
- Significant contributions of GridForce

Software Trends



Grid Technology

- Emerging enabling technology.
- Natural evolution of distributed systems and the Internet.
- Middleware supporting network of systems to facilitate sharing, standardization and openness.
- Infrastructure and application model dealing with sharing of compute cycles, data, storage and other resources.
- Promoted by NSF through its Network Middleware Initiative (NMI version 4).
- Publicized by prominent industries as on-demand computing, utility computing, etc.
- Move towards delivering "computing" to masses similar to other utilities (electricity and voice communication).

Adaptation of Grid Technology to CS-Curriculum

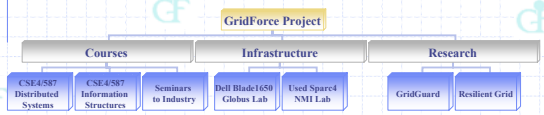
- Introduce grid technology into the CS undergraduate curriculum.
  - Goal: Design and deploy grid services and applications. Study grid application models.
  - Focus on lab exercises to illustrate fundamental grid concepts, and development of grid services and applications.
- Conduct seminars to industry.
  - Goal: Overview of grid technology landscape and its alignment to common technologies and application models.
  - Examine case-studies to expose potential uses of grid.
- Introduce grid to potential users of grid
  - Goal: Publicize the usage models of grid.
  - Use grid infrastructure for entry level courses in Sciences and Engineering.

GridForce

- Our adaptation is evolving into a comprehensive framework we call GridForce (**Grid For Collaboration and Education**):
  - Course curriculum,
  - Laboratory exercises (labs),
  - Infrastructure to support labs,
  - Research projects,
  - Industrial outreach.

GridForce Project Framework

GridForce is a comprehensive framework to adapt grid Computing into undergraduate curriculum.



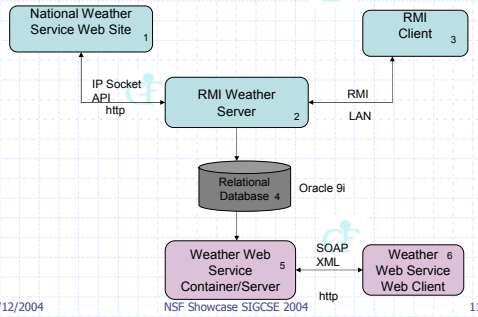
Courses: CSE4/586 Distributed Systems

- Learning outcome: fundamental concepts of distributed systems and grid.
- Lab exercises to support concepts:
  - Three-tier client server system using Web Services.
  - A simple grid framework.
  - Design and implementation of a grid service.
- Text: Distributed Systems: Concepts and Design (3rd Edition) by George Coulouris, Jean Dollimore, Tim Kindberg, Addison-Wesley Inc., 2000.
- Prerequisites: Data structures and algorithms, object-oriented design and development, working knowledge of Java.

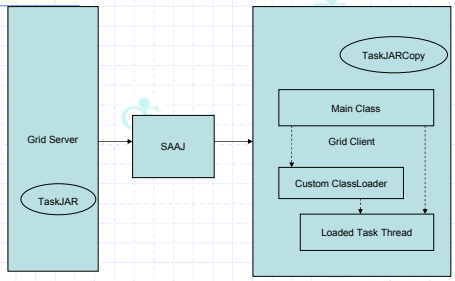
Courses: CSE 4/586: Lab Exercises

Exercise	Topic	Learning Objective
Lab1	Webservices	To understand the alignment of the grid technology to Web Services, WS Definition Language (WSDL) and service description using WSDL.
Lab2	Grid Infrastructure	A Webservices based grid.
Lab3	Grid Programming	Design and implement a grid-base service using Globus 3.0.2

Lab1: Web Services

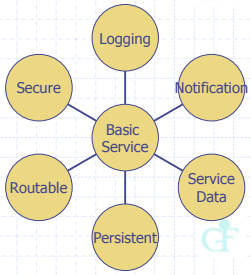


Lab2: Web Services-based Grid



Based on A simple do-it-yourself framework for grid computing by Anthony Karre

Lab3: Grid Service using Globus



Courses: CSE4/587:Information Structures

- Learning outcome: understand grid infrastructure and grid architecture, design and deploy grid services and grid applications.
- Lab exercises support:
  - Enterprise application using Java 2 Enterprise Edition.
  - Grid application in high performance area.
  - Service-oriented grid application.
- Text: "The Grid 2: Blueprint for a New Computing Infrastructure" by Ian Foster , Carl Kesselman, Morgan-Kauffmann, 2004.
- Prerequisites: Data structures and algorithms, object-oriented design and development, working knowledge of Java, fundamentals of client/server architectures.

Courses: CSE4/587: Lab Exercises

Exercise	Topic	Learning Objective
Lab1	Commercial Application	Study requirements of a commercial domain and implement an application.
Lab2	High performance Application.	Study requirements of scientific/business domain and implement compute intensive application.
Lab3	Defining a high-level grid service	Workflow service, a business process, improvements to QoS

Fundamental Grid Concepts Covered

- N-tier client server system
- Web applications
- Component programming
- Grid service
- Open grid services architecture (OGSA)
- Open grid services infrastructure (OGSI)
- Logging, notification and service data
- Virtualization, federation, provisioning.

Fundamental Knowledge Areas

- From ACM Curricula 2001:
- NC1: Net-centric computing: Distributed Systems
- NC5: Building web applications
- SE2: Using APIs
- SE3: Software tools and environments
- SE9: Component-based Programming
- SE12: Specialized system development
- CN4: High Performance computing

Tools and Technologies covered

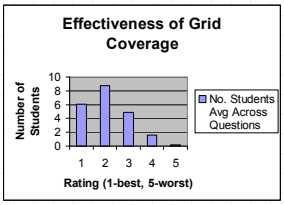
- Technologies include:
  - XML and SOAP
  - Web services (service definition, implementation and deployment)
  - Java 2 Enterprise Edition (Enterprise Java Beans)
  - Globus Toolkit 3.0.2 (GT3)
- Tools include:
  - UML (Unified Modeling Language) for design representation
  - Apache Ant: XML-based build tool

Outcome Assessment

- End of the course questionnaire is used to assess the effectiveness of the courses.
  - prepared by an external consultant (Dr. Neal of Erie Community College)
- Mainly multiple choice questions with a few short answer questions.
- The overall effectiveness of the CSE4/586 course as measured by the average of ratings for the 42 questions. (1- best to 5-worst) is shown.

Effectiveness of Adaptation (CSE486/586)

- Survey with 42 multiple choice questions pertaining to coverage of grid in CSE4/586.
- Average rating among 20 students who took the survey is shown.

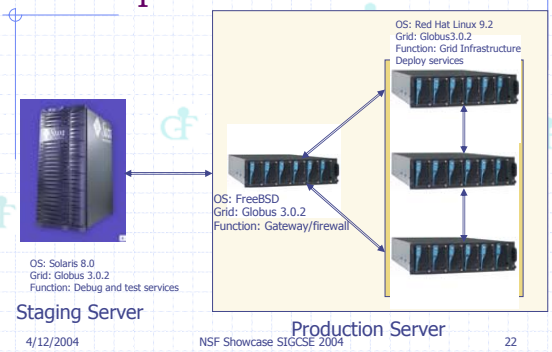


- External evaluator identified 7 areas for improvement.
  - Two of these pointed to unavailability of grid programming infrastructure for students to use.
  - We have remedied this situation with more than one grid lab infrastructure.

Infrastructure: LinuxGrid

- Goal: To facilitate development of service-oriented applications for the grid.
- Two major components: Staging server and Production grid Server.
- Grid application are developed and tested on staging server and deployed on a production server.
- Production grid server:
  - Three compute nodes with Red Hat Linux and Globus 3.0.2 instance.
  - One utility gateway node with Free BSD and Globus 3.0.2.
- Lab 1 will be deployed on the staging server, Lab 3 on the production grid.

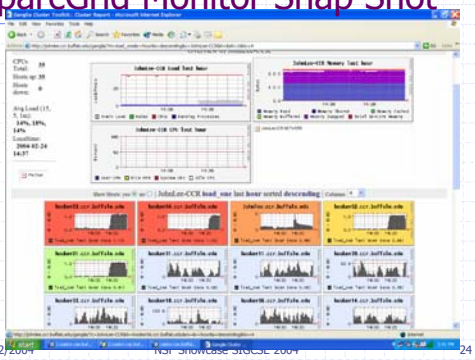
Development Environment



Infrastructure: SparcGrid

- Goal: To run jobs submitted in a distributed manner on a Condor-based computational cluster Condor.
- Composed of 50 Sun recyclable used Sparc4 machines, which form computational nodes, headed by a front-end Sun server.
- The installation scripts are custom-written facilitating running of jobs in a distributed manner.
- Partially supported by Center for Computational Research (CCR).
- Lab2 will be developed, deployed and tested on this infrastructure.

SparcGrid Monitor Snap Shot



Industrial Training Tier

- In collaboration with The Center for Industrial Effectiveness (TCIE) of University at Buffalo (UB).
- A Two-hour breakfast seminar introducing grid technology to business decision makers and potential adopters.
- Topics include:
  - Grid application domains
  - Grid application models
  - How relate grid to currently used technologies and
  - Details of the grid infrastructures currently operational at UB.

Resources for Adoption

- GridForce is modular that all or parts of it can be adopted by educators and practitioners.
- Course curriculum, project descriptions, solutions and lecture material are available online at www.cse.buffalo.edu/gridforce

Challenges in Adapting Grid Technology

- Adding to existing curriculum.
  - Solution: Addressed through labs.
- Adapting to versions of software and toolkits.
  - Solution: Discusses differences; however work with the latest version.
- Managing students with deficiencies in their technical background.
  - Solution: Special coverage during recitations.
- Maintaining grid infrastructure for hands-on labs.
  - Solution: Include a system administration support person in the implementation of curriculum.
- Lack of appropriate text books:
  - Solution: Good area for anybody with expertise to write a book.

Contributions

- Expected number of students *directly* impacted: 200+ per year. With proper dissemination this will be much higher.
- Comprehensive framework covering grid technology in course curriculum, lab exercises, infrastructure to support labs, and applied research.
- Coverage addressing needs at various levels: undergraduate, graduate to industrial workforce and decision makers.
- Offers a model for adaptation of ever changing technology landscape.