
The Grid:

Will It Live Up to the Hype?

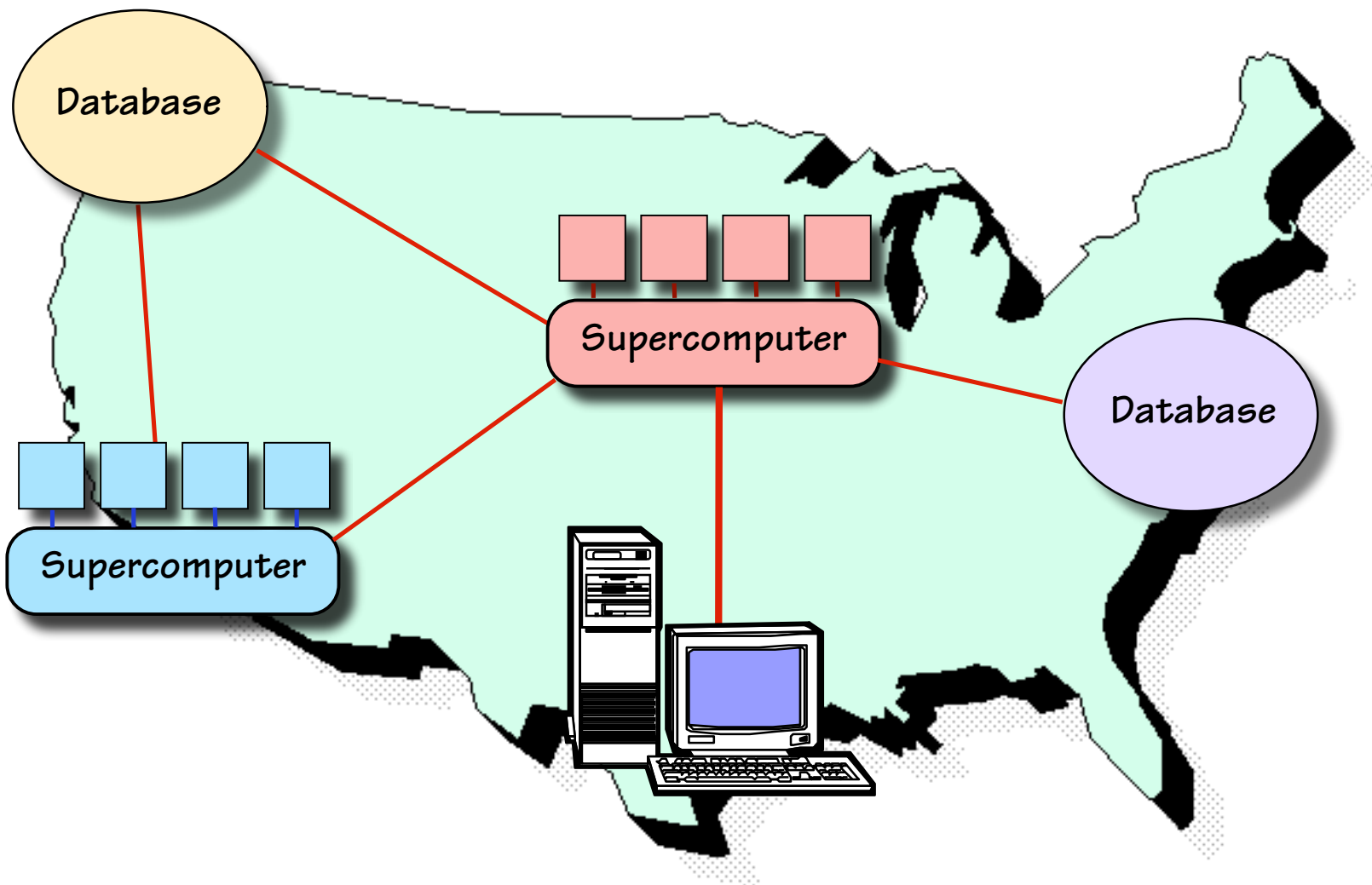
Ken Kennedy
Center for High Performance Software
Rice University

<http://cs.rice.edu/~ken/Presentations/GridVentureForum05.pdf>

The Grid: Hype

- **Global Computing Resources As the Ultimate Supercomputer**
 - Millions of idle processors waiting to solve your problem
 - All the worlds databases and scientific instruments available as peripheral devices
- **Computing for Free**
 - Exploiting idle cycles means less overall expense to society
- **Easy Problem-Solving**
 - Launch a problem and let the Grid solve it for you
 - Do for problem-solving what the Internet has done for information access

The Grid: A Global Problem-Solving System



The Grid: Reality

- **Some Interesting and Useful Grid Applications**
 - SETI@Home, scientific portals
- **Grid Application Development Is Possible**
 - Commercial services model: Grid services like web services
 - Globus: Resource location, distributed application launch
 - Condor, DAGMan: Simple workflow management tools
- **But Hard**
 - Scheduling applications by hand onto millions of resources
 - Adapting to dynamically changing resources (and faults!)
 - Accounting and security
- **Challenges**
 - Broaden the community of users
 - If programming is hard, the Grid will not reach its potential
 - Manage the divergence of commercial and scientific computing

Our Approach

- Virtual Grid Application Development Software (VGrADS) Project
 - NSF Information Technology Research (ITR) Large Award
 - 7 institutions, 12 principal investigators
- What do we propose as a solution?
 - Separate application development from resource management
 - Through an abstraction called the Virtual Grid
 - Provide tools to bridge the gap between conventional and Grid computation
 - Scheduling, resource management, distributed launch, simple programming models, grid economies

The VGrADS Team

- VGrADS is an NSF-funded Information Technology Research project



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Dan Reed



RICE

Keith Cooper
Ken Kennedy
Charles Koelbel
Richard Tapia
Linda Torczon



Jack Dongarra



Carl Kesselman



Fran Berman
Andrew Chien
Henri Casanova



Rich Wolski



Lennart Johnson

- Plus many graduate students, postdocs, and technical staff!

VGrADS Project Vision

- Virtual Grid Abstraction
 - Separation of Concerns
 - Resource location, reservation, and management
 - Application development and resource requirement specification
 - Permits true scalability and control of resources
- Tools for Application Development
 - Abstract programming interfaces
 - Easy application scheduling, launch, and management
 - Workflow graphs and tightly-coupled computations
- Support for Fault Tolerance and Rescheduling/Migration
 - Collaboration between application and virtual grid execution system
- Research Driven by Real Application Needs
 - EMAN, LEAD, GridSAT, EOL

Programming Tools

Focus: Automating critical application-development steps:

- **Initiating and managing application execution**
 - Optimize and launch application on heterogeneous resources
 - Support for fault tolerance and rescheduling/migration
- **Scheduling application workflows**
 - Whole-workflow scheduling using performance models
 - Two-phase scheduling using Virtual Grid prescreening
 - Scheduling onto resources with batch queues (no reservations)
- **Constructing performance models**
 - Automatically from application binaries
 - Cross-platform modeling
- **Building workflow graphs from high-level scripts**
 - Examples: Python (EMAN), OGRE (LEAD), Matlab

VGrADS Applications

Philosophy: Computer Science Research Driven by Applications

- **EMAN**
 - Construction of 3-D models from 2-D electron micrographs
 - Collaboration with National Center for Macromolecular Imaging
 - Wah Chiu, Baylor College of Medicine
- **LEAD**
 - Collection of applications for real-time weather prediction, sponsored by NSF ITR Large
 - Kelvin Droegemeier, University of Oklahoma
- **GridSAT**
 - Satisfiability on the Grid
 - World's fastest implementation

Leverage

- **National Middleware Initiative**
 - Community infrastructure that can be used as a base
 - Vehicle for deployment of successful research
- **State of Texas Support**
 - **LEARN:**
 - Support for statewide networking
 - 33 institutions \$7.5 million over two years
 - **TIGRE**
 - Application driven software stack
 - \$2.5 million for two years shared by 5 institutions
 - Rice and UH both participants
- **Houston Research and Education Network**
 - Rice, UH, Texas Med Center, National Lambda Rail
 - Abovenet contribution \$3.4 million



Summary

- **The Grid:**
 - Lots of Hype
 - Lots of Potential
- **The Problem**
 - Application development, scheduling, and launch are hard
 - Standard OS functionality is missing
 - e.g., accounting, security, rescheduling/migration, fault tolerance
- **The VGrADS Project**
 - Construct software tools based on separation of concerns
 - Allow developers to focus on their applications
 - Drive research with real applications of importance to the nation
 - Contribute to state and national infrastructure projects and testbeds
 - TIGRE, LEARN, NSF TeraGrid