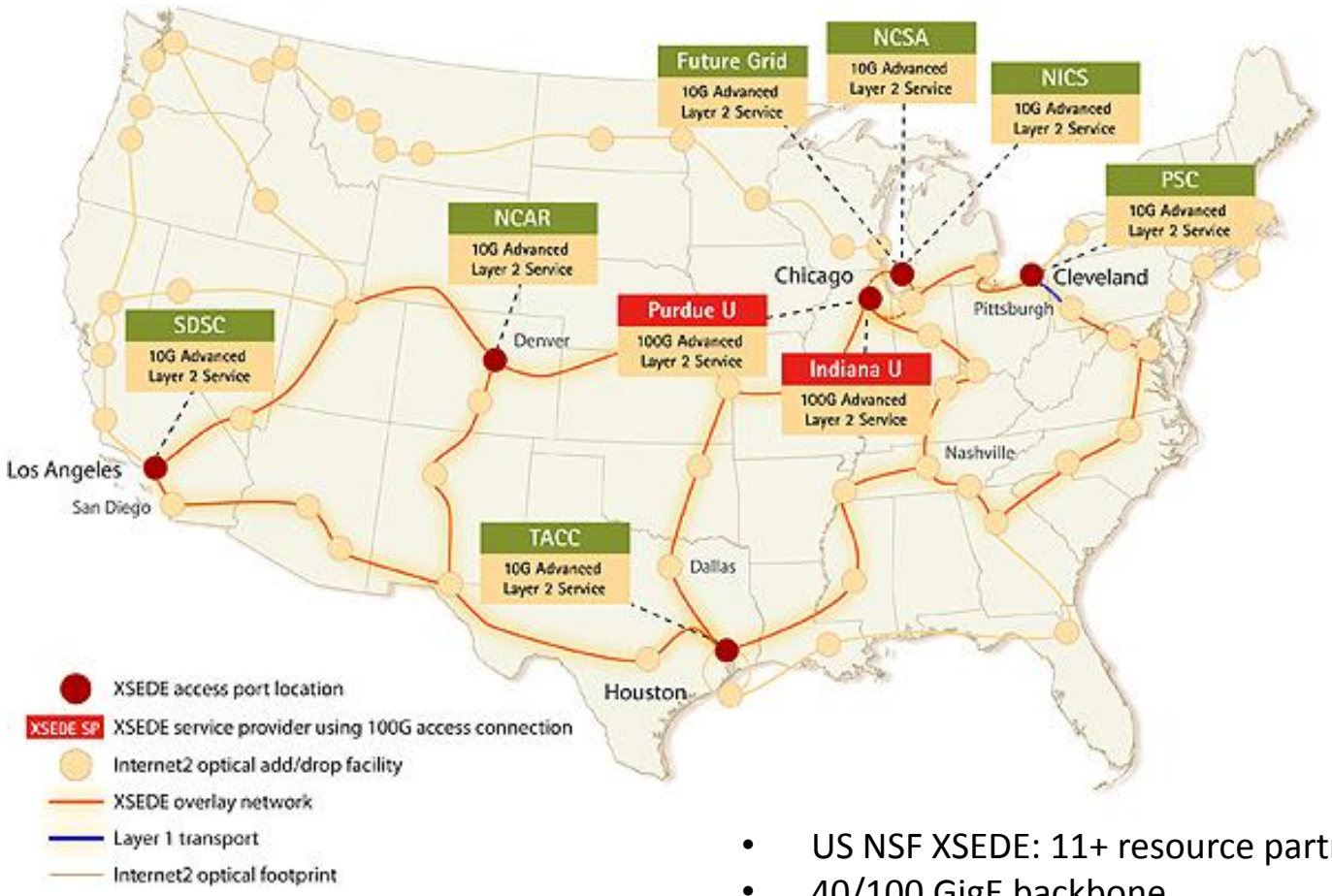
Integrating HPC Resources, Services, and Cyberinfrastructure to Develop Science Applications

Mary P. Thomas

Department of Computer Science, and
Computational Sciences Research Center, San Diego State University

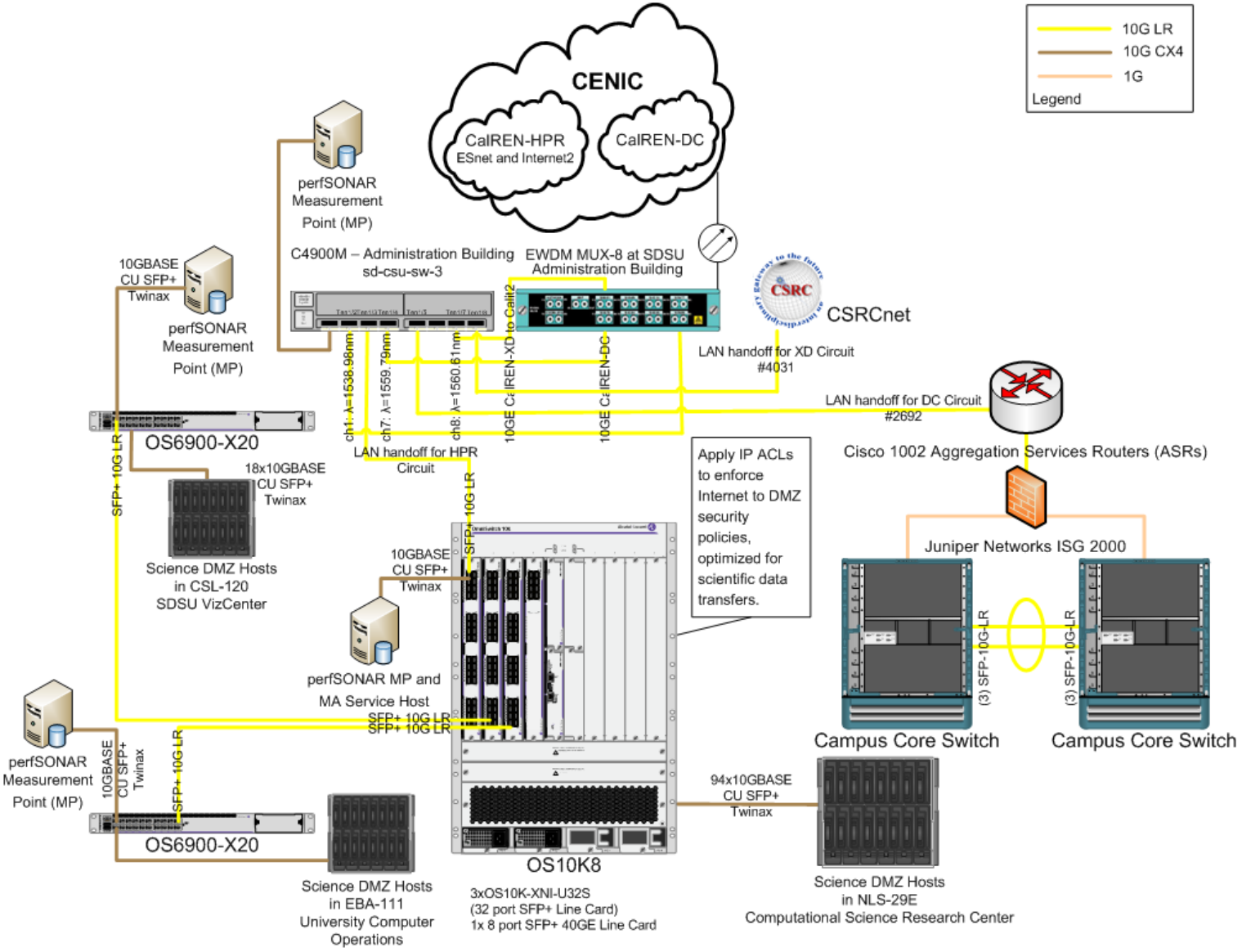
Human Dynamics Lightning Talks
San Diego State University
April 24, 2014

NSF XSEDE: An Example of the Scale of the Challenge



- US NSF XSEDE: 11+ resource partners
- 40/100 GigE backbone
- PetaFlops (aggregate), Multi PByte storage
- Gateways: 100's of large projects, 1000's of users
- Future Grid Project: cloud computing resources

Smaller Scale CI: SDSU Science DMZ

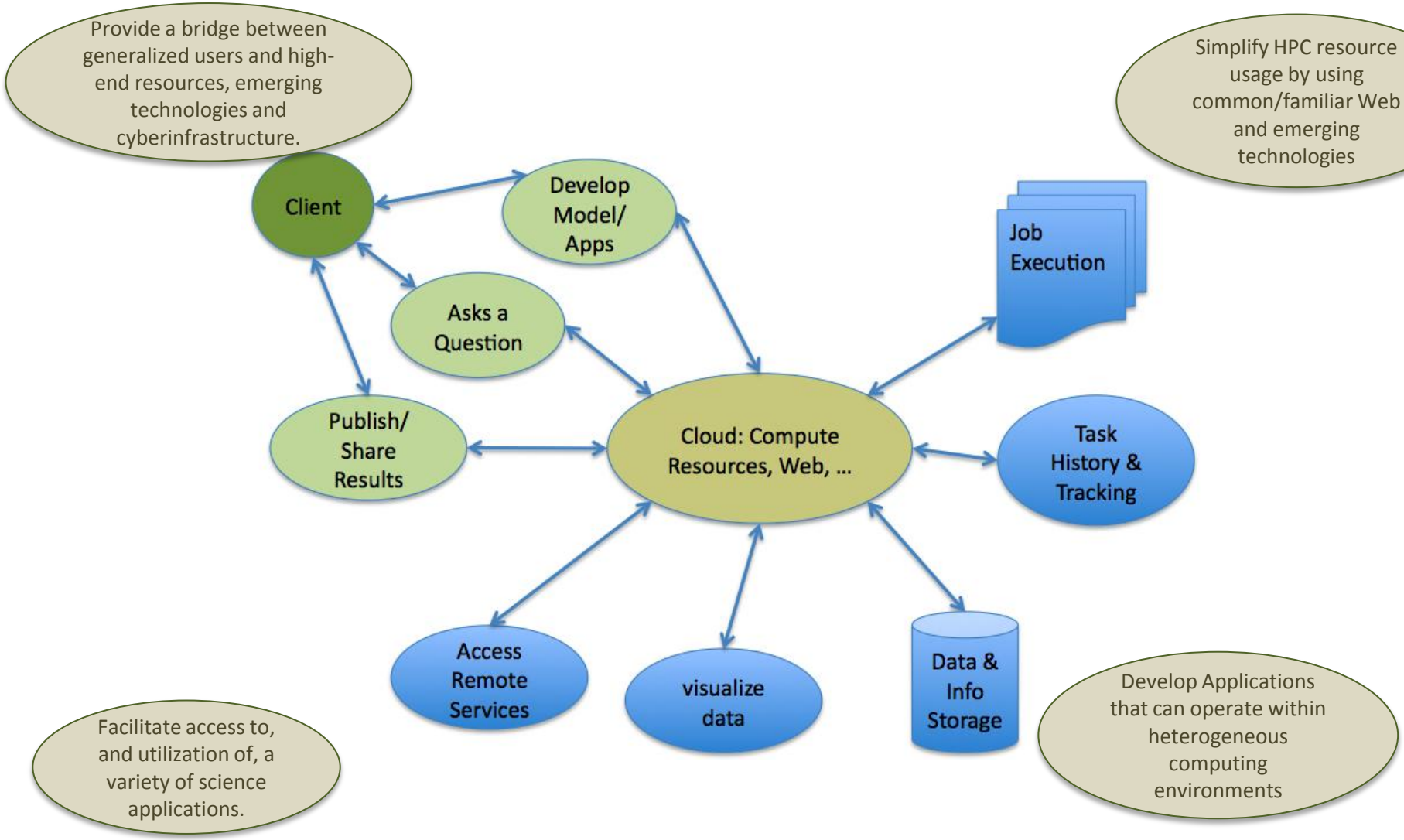


NSF Office of CyberInfrastructure CC-NIE Grant 1245312

The SDSU Cyberinfrastructure Web Application Framework (CyberWeb)

- CyberWeb simplifies the utilization of heterogeneous, computational environments required by high-performance computing applications
- Part of an ongoing NSF project, the Open Grid Computing Environments (OGCE) project, which has a focus on XSEDE Gateway projects
  - Evolved from Grid Portal (GridPort) Toolkit Project
- Team effort:
  - PI: Mary Thomas (CS & CSRC Departments)
  - M.S. students Hetang Shah, Smita More, and Carny Cheng working on distinctly different aspects

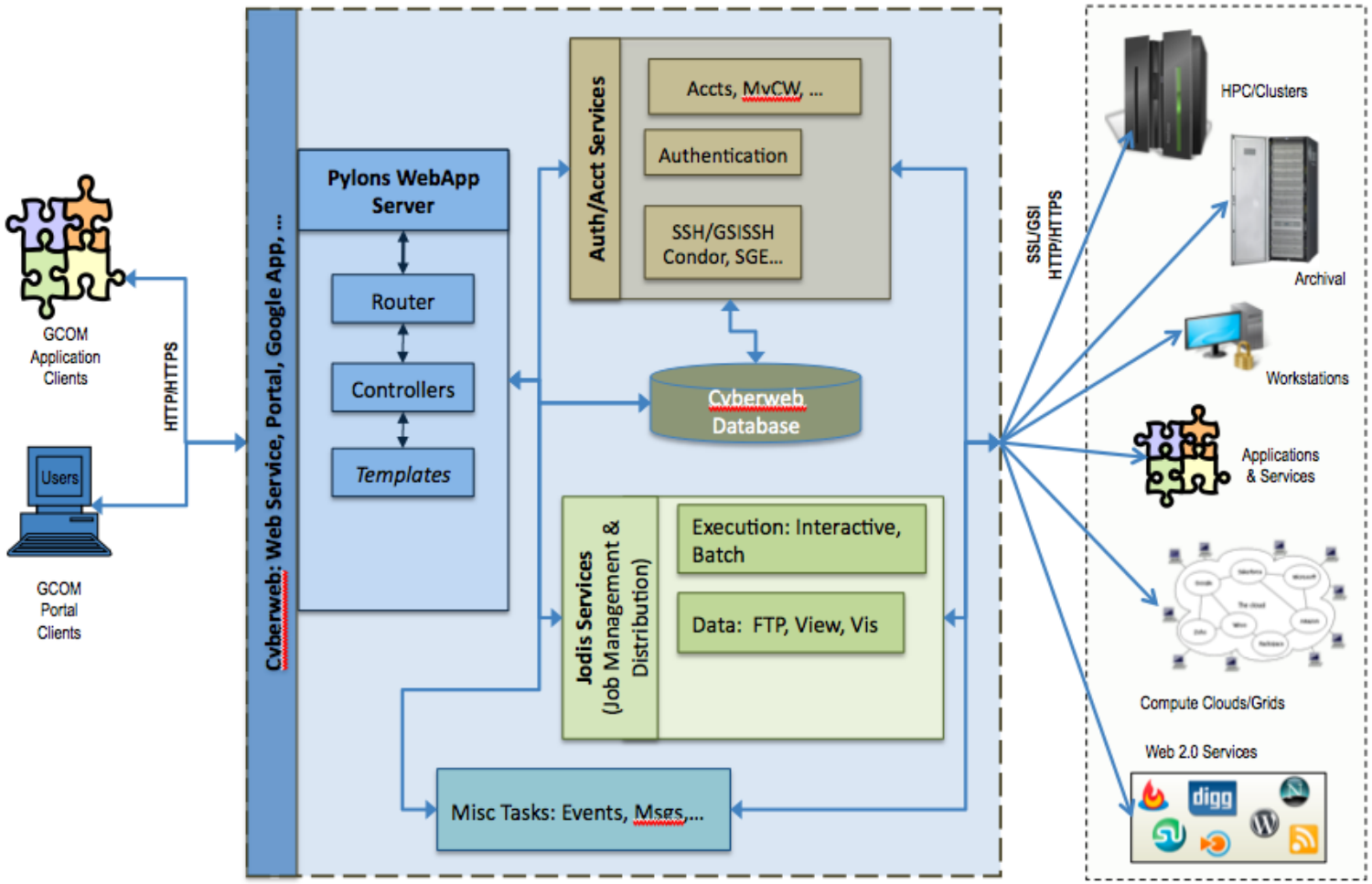
CyberWeb: Support Computational Environments



CyberWeb Architecture & Technologies: Classic 3-Tier Design

- Clients:
  - remote applications, Web services
  - Web portals.
  - Command line interface (CLI)
- Services Oriented Architecture:
  - capable of hosting/exposing any functionality as a service.
- Web Service Gateway Interface (WSGI):
  - Pylons: Web 2.0 WSGI application framework
  - Relational databases
  - XML, JavaScript, AJAX,
  - Google Gadgets, social networks
- Security & Authentication
  - Automatic and integrated
  - Support multiple protocols
- Dynamic Database:
  - admin Web pages, for configuring CyberWeb installations, applications, users, remote resources and services.
- Job Execution:
  - Job distribution Web service framework for task execution and management.
- Data Management:
  - distributed
- Connectivity:
  - Heterogeneous resources and services (remote or local).
  - Any network (TCP/IP, 10GigE)

CyberWeb Architecture



7

CyberWeb: Features & Capabilities

- Dynamic Database
  - core to everything
- Security/Authentication:
  - HTTPS/SSH/GSI Users
  - Access Control List (ACL)
  - Validated authentication information visible to all modules and components being used by the portal.
  - Map CyberWeb users to accounts on remote resources
- Supports multiple applications:
  - Configurable in DB
- Heterogeneous:
  - Accesses multiple hosts and queuing systems
  - Distributed environment
  - New systems added via database admin
- Job Execution:
  - Dynamically build & deploy jobs
  - Moves and stages I/O files
  - Interactive unix cmds
  - Queing/batch jobs
  - Job Monitoring
- Data Management:
  - 3rd party file transfer
  - Job staging & history
  - “Plug-n-play” approach for resource configuration and its use
- Post Processing Services:
  - Visualization, Analysis

CyberWeb Database

- Implemented using
  - Pylons/SQLAlchemy (API to multiple databases)
  - MySQL and SQLite + JSON
- Design based on major existing grid databases
  - TeraGrid; Open Grid Forum; FutureGrid/Cyberaide
- Database is core to everything:
  - Resource configuration (add/remove hosts, queues,
  - Accounts: map CW account to users remote host account on TG, workstations, Condor, SGE, etc.
  - Authentication: ssh, gsi, condor, srb, sge, …
  - File management; input/output; move files
  - Dynamic project/task naming

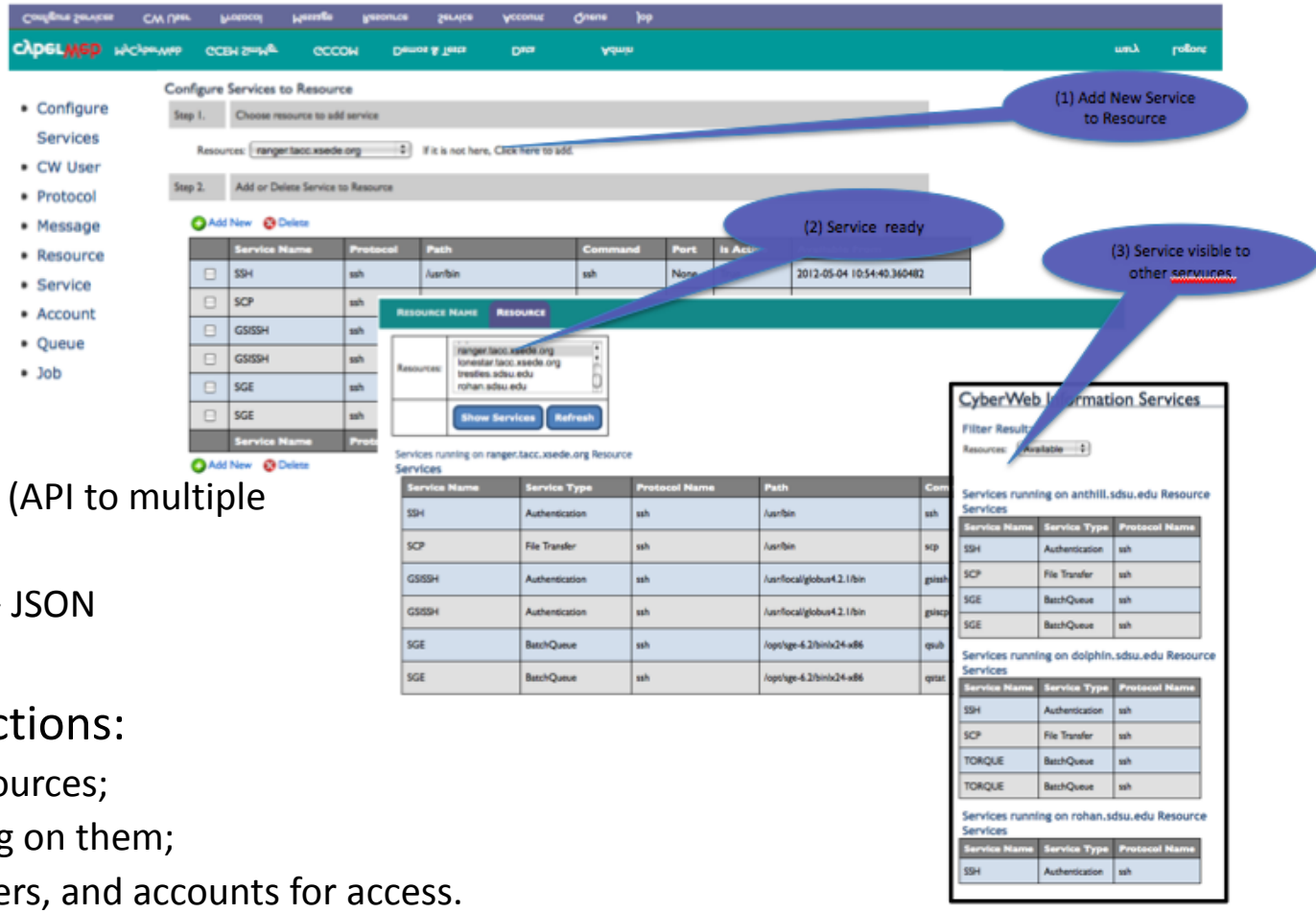
CyberWeb: Database Admin

- Design based on existing grid RDB's:
  - XSEDE/TeraGrid
  - Open Grid Forum FutureGrid
  - Cyberaide
- Technologies:
  - Pylons/SQLAlchemy (API to multiple databases)
  - MySQL and SQLite + JSON
  - JavaScript and Ajax
- Dynamic admin functions:

  ①add and configure resources;

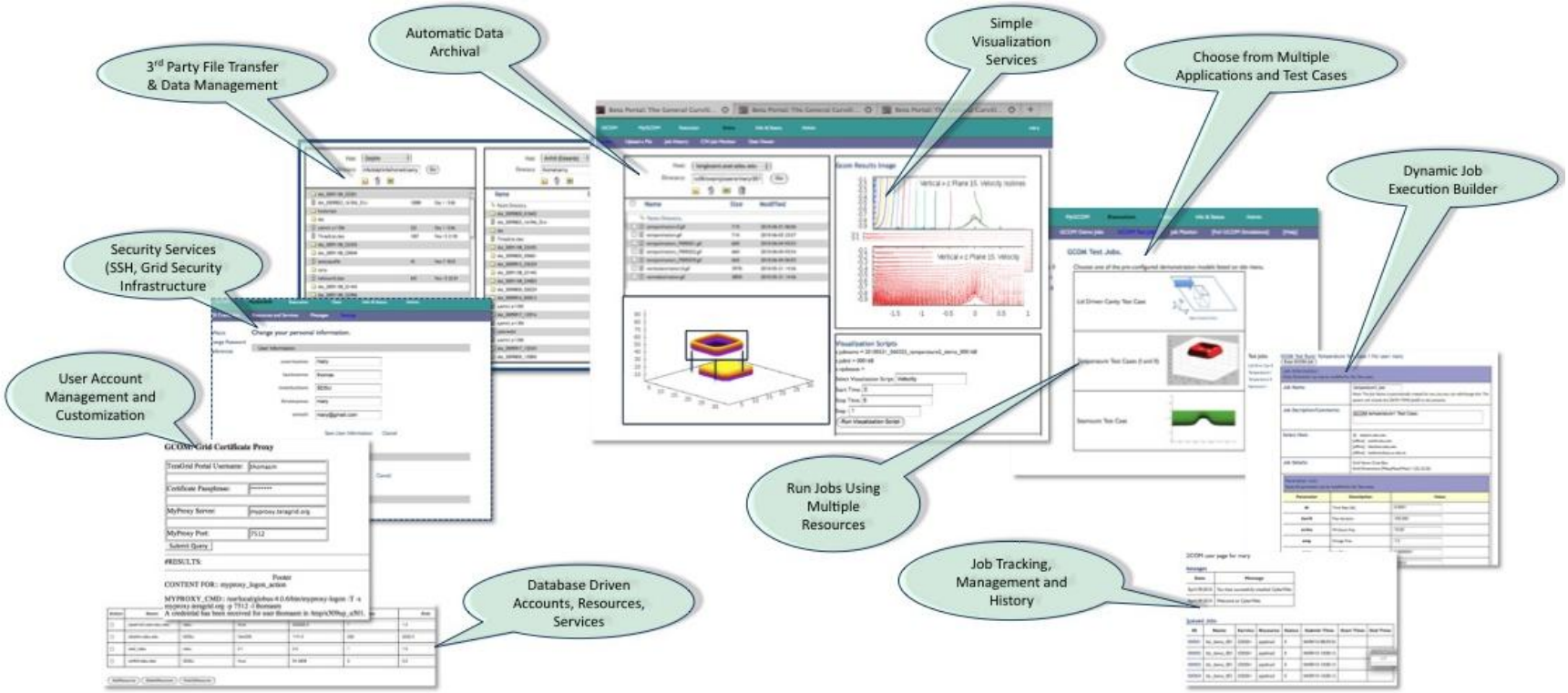
  ②define services running on them;

  ③create and validate users, and accounts for access.



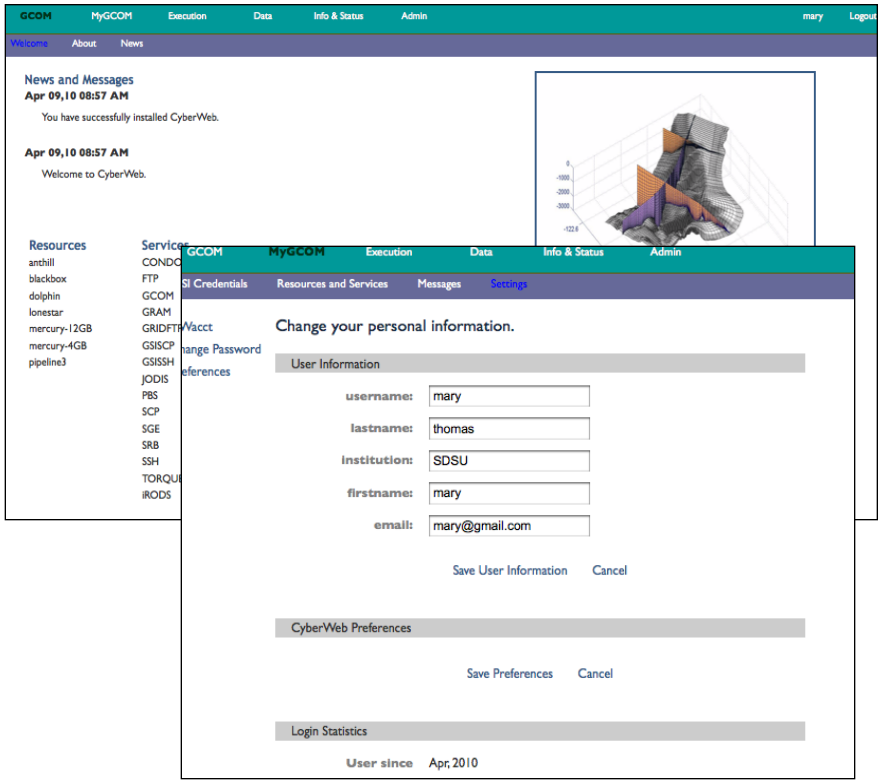
- "Live" machines: available for use immediately by other services.

CyberWeb App: GCEM Coastal Simulation Portal



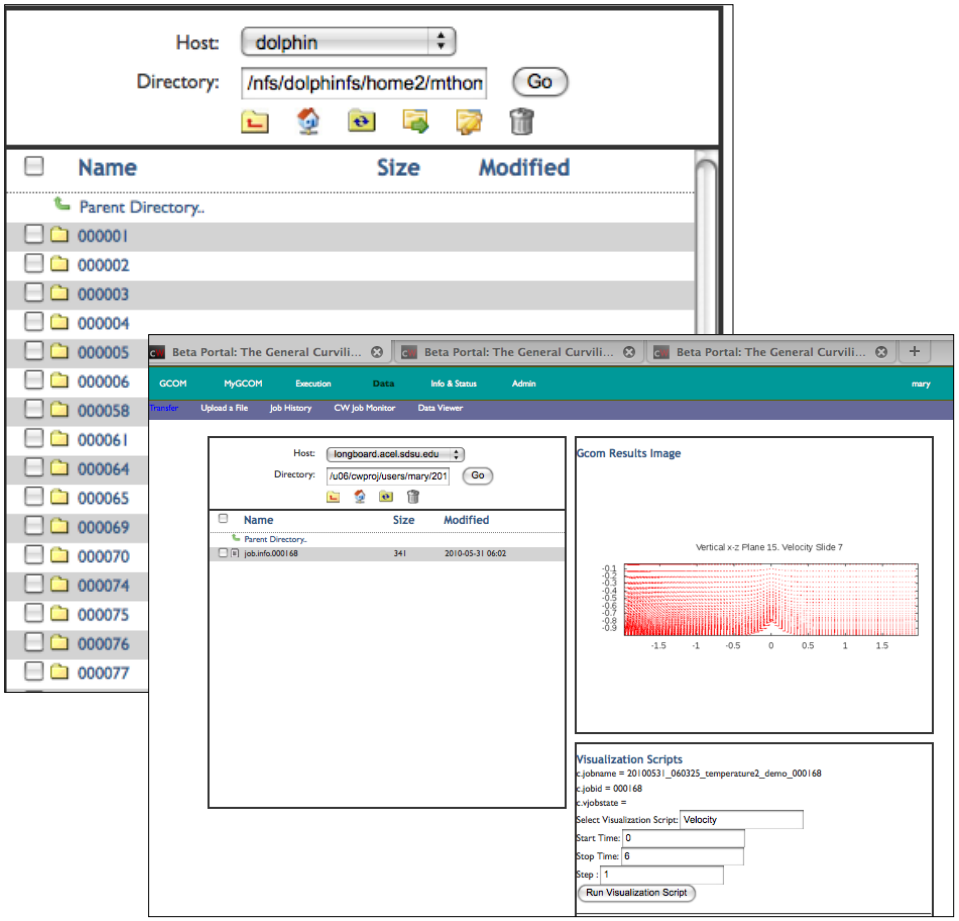
CyberWeb: Home Page

- Customized Home Page
- Dynamic view of available resources and services
  - Tested by job distribution service
- News/Events/Messaging
- Account creation, customize, set preferences
- HTTPS



GCEM Portal: File Transfer and Data Management

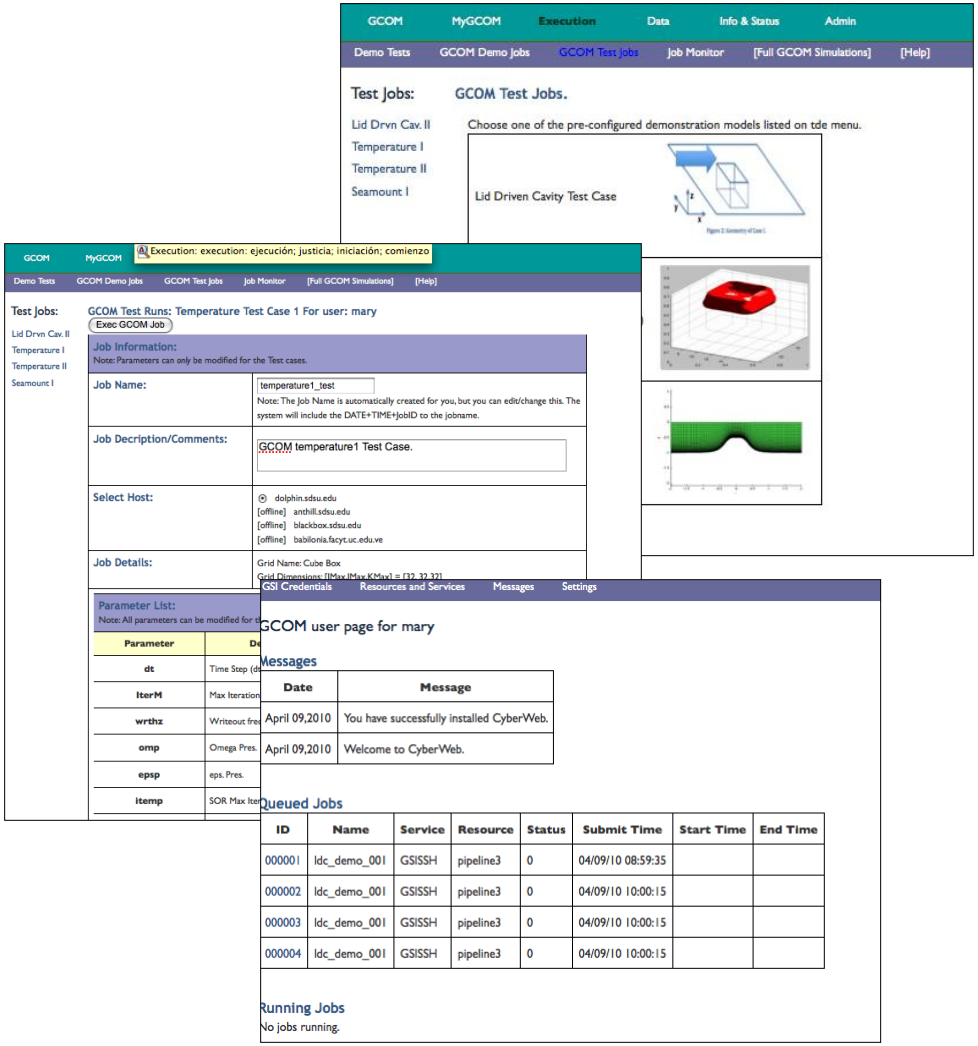
- Big Data Support: working with GlobusOnline project (paper at XSEDE'13*)
- 3rd party file transfer between resources, local host.
- View file contents
- Quick viz views



XSEDE'13 Conference, July 20-24, San Diego, CA. , http://www.xsede.org

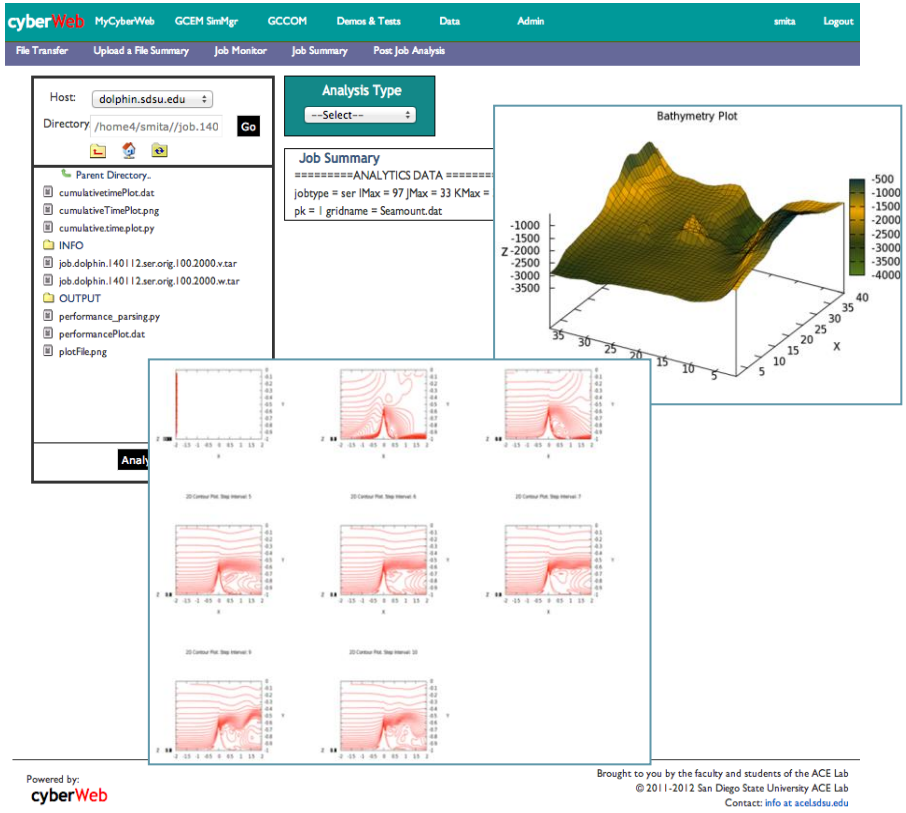
CyberWeb: Job Management

- Select Jobs:
  - Pulled from DB
- Build and submit job
  - “Live” resources pulled from db
  - store in historical database
- Monitor jobs and status
- Resources & Services monitor



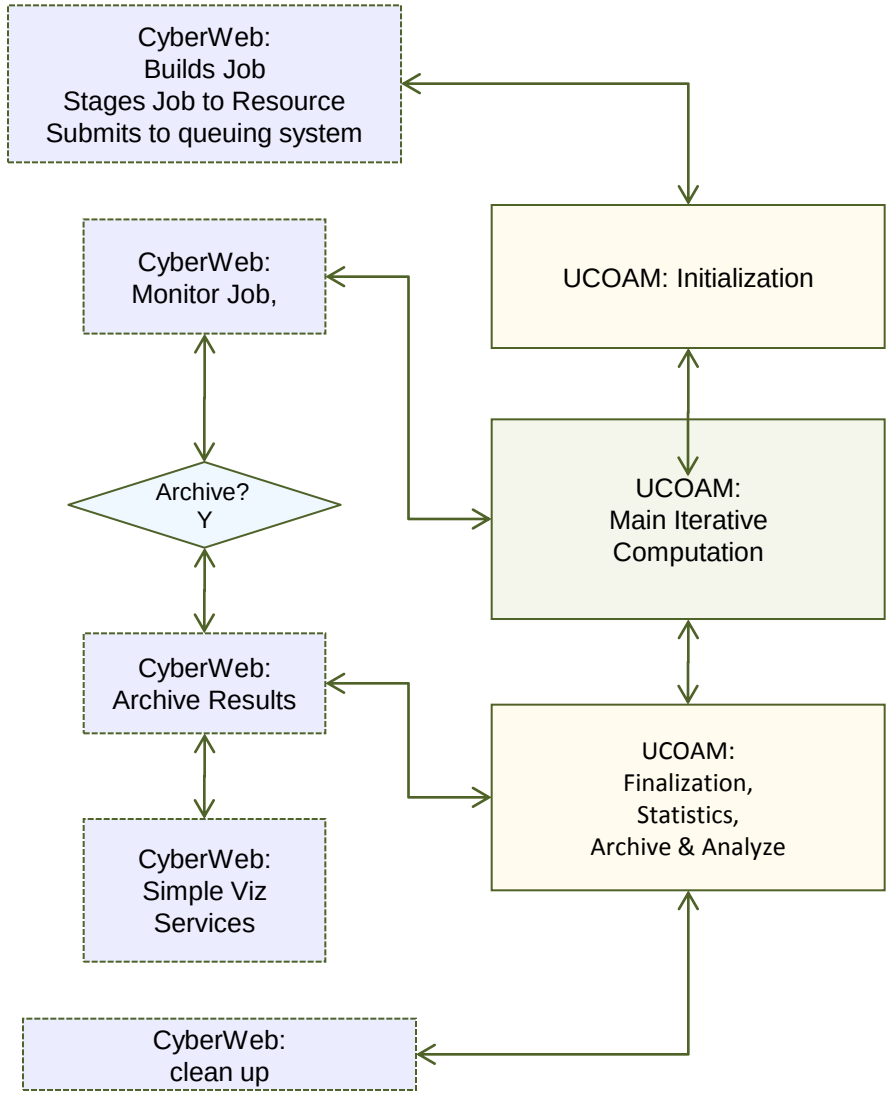
CyberWeb Visualization (CyberViz)

- Based on Python & Gnuplot lib
- View interim or final data
- Set parameters such as job, plot type, scale
- Requests trigger data transfer from compute host to archival system
- Generate images or movies



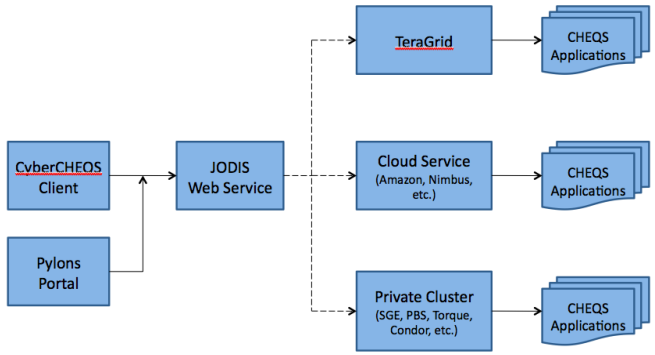
CyberWeb App: Running Parallel UCOAM Applications

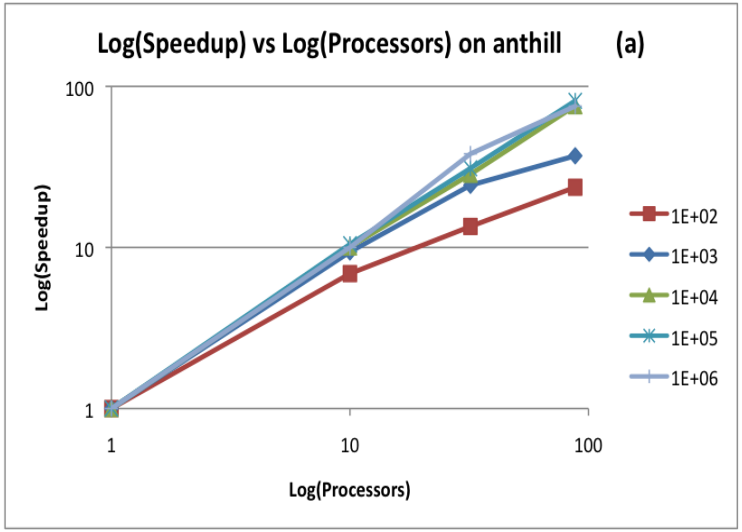
- parUCOAM code integrated in CE
- Application deployed to remote resources
- CA/CyberWeb
  - builds and deploys test cases
  - Manages results
  - Performs simple visualization



CyberWeb App: CyberCHEQS Project

- Combustion simulations to model chemical reacive flows.
- Uses Web service to update chemical composition
- CyberWeb services used to run jobs in parallel
- Scales to millions of jobs on hundreds of node - EP
- Collaboration with S. Bhattacharji and C. Paolini (SDSU)





Future Work

- Job execution/task management:
  - History
  - Job builder & compiler
  - Auto scheduler, interrupt/kill/steering
- Data & visualization
  - Access to larger archival resources
  - “Big Data:” Integrate GlobusOnline & other tools
- Automate application deployment
- Expand post processing capabilities
- Add Cloud resources
- Simplify installation
  - Unit Test System
  - Installation via Python egg
- Develop large scale gateways on XSEDE

Thank You

- Questions:
  - mthomas@mail.sdsu.edu
  - http://acel.sdsu.edu
- References:

  - M P Thomas, C Cheng, S More, and H Shah. Integrating HPC Resources , Services , and Cyberinfrastructure to Develop Science Applications Using Web Application Frameworks. In *Int. Conf. Parallel Distrib. Process. Tech. Appl.*, volume 2, pages 421–427, Las Vegas, NV, 2012.
  - Mary Thomas and Jose Castillo. A Cyberinfrastructure-Based Computational Environment for Unified Curvilinear Ocean Atmospheric Model (UCOAM). In *4th Int. Congr. Comput. Eng. Sci.*, 2013.
  - [21] Mary P Thomas and Jose E Castillo. Development of a Computational Environment for the General Curvilinear Ocean Model (GCOM). *J. Phys. Conf. Ser.*, 180(1), 2012.