

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

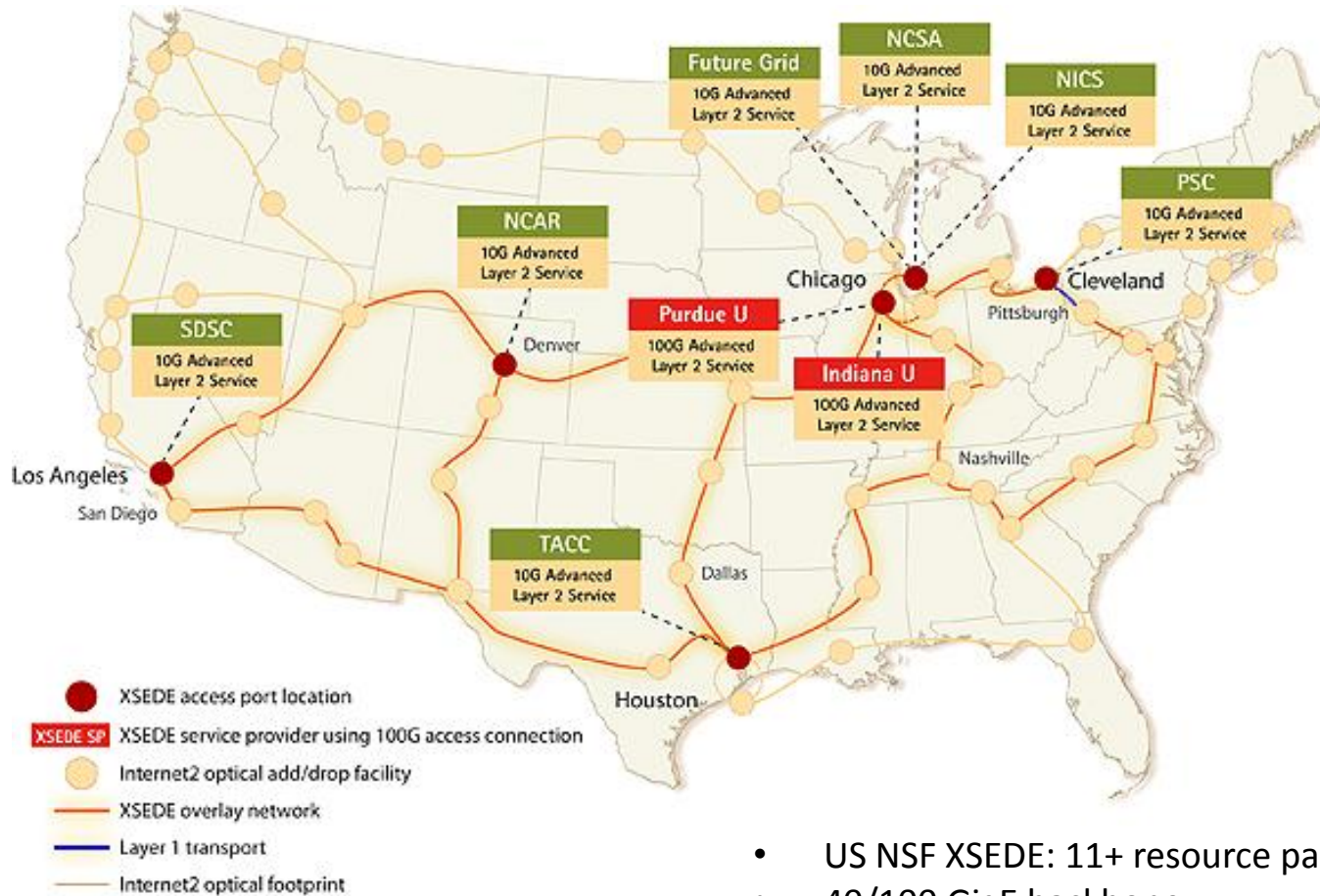
Mary P. Thomas

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

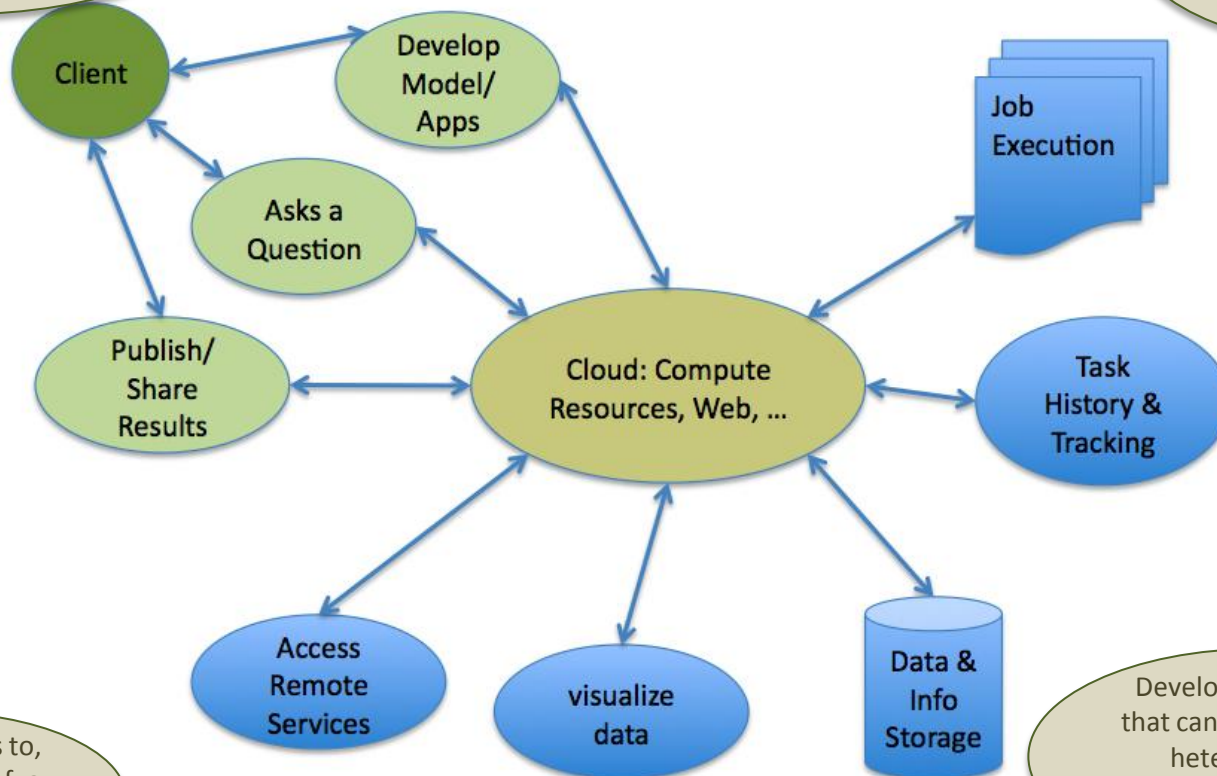
The SDSU Cyberinfrastructure Web Application Framework (CyberWeb)

- 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

Provide a bridge between generalized users and high-end resources, emerging technologies and cyberinfrastructure.

Simplify HPC resource usage by using common/familiar Web and emerging technologies



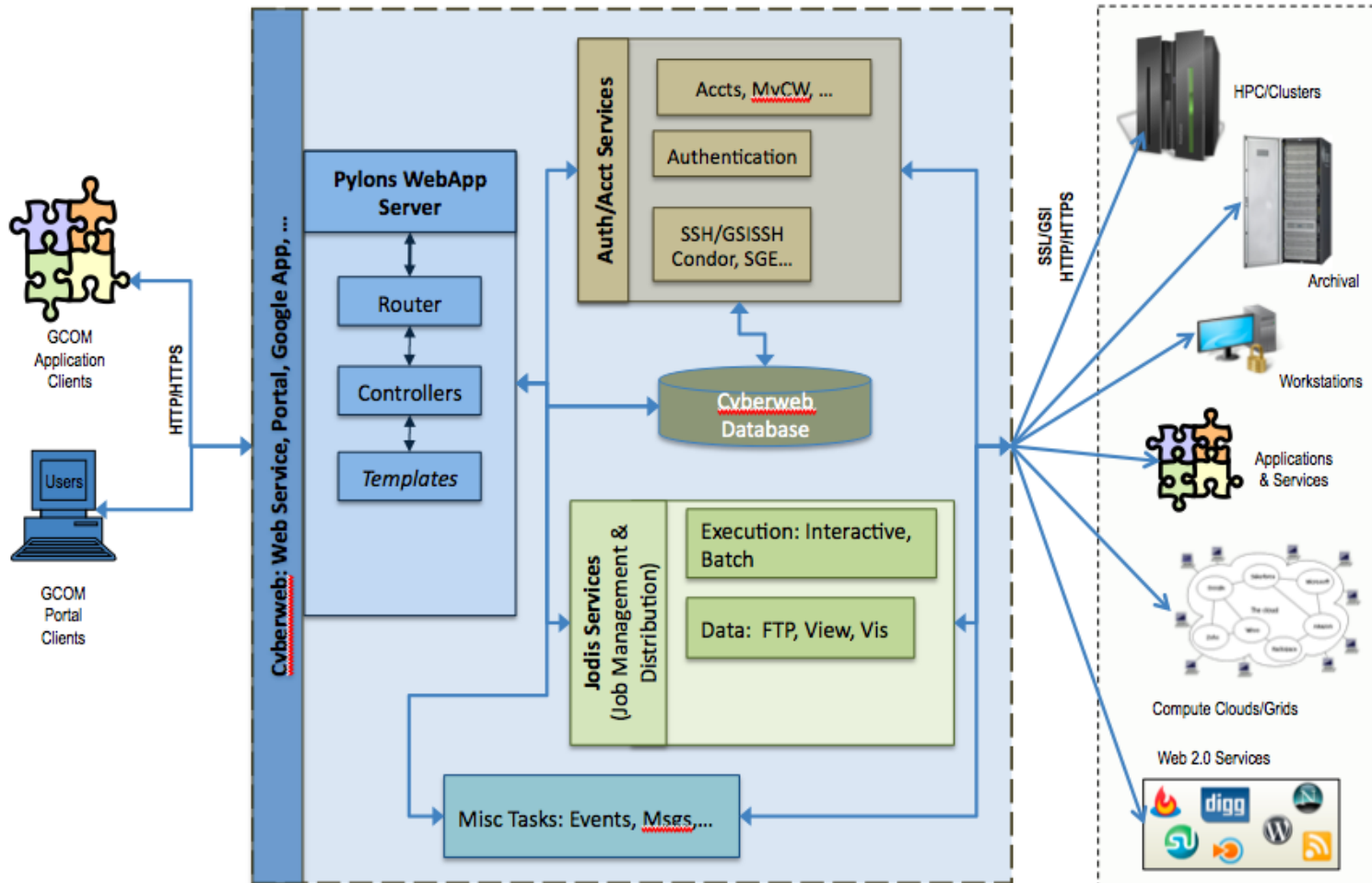
Facilitate access to, and utilization of, a variety of science applications.

Develop Applications that can operate within heterogeneous computing 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



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:

- ① add and configure resources;
- ② define services running on them;
- ③ create and validate users, and accounts for access.

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

The screenshot displays the 'Configure Services to Resource' interface. It is divided into two main steps:

Step 1: Choose resource to add service
 Resources: ranger.tacc.xsede.org (If it is not here, Click here to add)

Step 2: Add or Delete Service to Resource
 Add New Delete

A table lists services with columns: Service Name, Protocol, Path, Command, Port, and Is Active. The table includes entries for SSH, SCP, GSSSH, and SGE.

A callout bubble (1) points to the 'Add New' button: (1) Add New Service to Resource

A callout bubble (2) points to the 'Show Services' button: (2) Service ready

A callout bubble (3) points to the 'Refresh' button: (3) Service visible to other services

Below the main interface, there are three panels showing 'Services running on' different resources:

- Services running on ranger.tacc.xsede.org Resource**

Service Name	Service Type	Protocol Name	Path	Com
SSH	Authentication	ssh	/usr/bin	ssh
SCP	File Transfer	ssh	/usr/bin	sftp
GSSSH	Authentication	ssh	/usr/local/globus4.2.1/bin	gssh
GSSSH	Authentication	ssh	/usr/local/globus4.2.1/bin	gscp
SGE	BatchQueue	ssh	/opt/lge-6.2/bin/x86	qsub
SGE	BatchQueue	ssh	/opt/lge-6.2/bin/x86	qstat
- Services running on anthill.sdsu.edu Resource**

Service Name	Service Type	Protocol Name
SSH	Authentication	ssh
SCP	File Transfer	ssh
SGE	BatchQueue	ssh
SGE	BatchQueue	ssh
- Services running on dolphin.sdsu.edu Resource**

Service Name	Service Type	Protocol Name
SSH	Authentication	ssh
SCP	File Transfer	ssh
TORQUE	BatchQueue	ssh
TORQUE	BatchQueue	ssh
- Services running on rohan.sdsu.edu Resource**

Service Name	Service Type	Protocol Name
SSH	Authentication	ssh

CyberWeb App: GCEM Coastal Simulation Portal

The image displays a collage of screenshots from the GCEM Coastal Simulation Portal, with callouts highlighting various features:

- 3rd Party File Transfer & Data Management**: A screenshot showing a file management interface with columns for Name, Size, and Modified.
- Automatic Data Archival**: A screenshot showing a list of simulation jobs with columns for Name, Size, and Modified.
- Simple Visualization Services**: A screenshot showing a 3D visualization of a coastal simulation, with a callout pointing to a graph titled "Vertical z-Plane (z, Velocity) (m/s)".
- Choose from Multiple Applications and Test Cases**: A screenshot showing a "GCEM Test Jobs" selection interface with options like "Let Ocean Only Test Case" and "Let Ocean & Land Test Case".
- Dynamic Job Execution Builder**: A screenshot showing a "GCEM Test Jobs" configuration interface with fields for "Job Name", "Job Description", and "Job Status".
- Run Jobs Using Multiple Resources**: A screenshot showing a "GCEM Test Jobs" configuration interface with a "Resources" table.
- Job Tracking, Management and History**: A screenshot showing a "GCEM Test Jobs" configuration interface with a "Job Status" table.
- Database Driven Accounts, Resources, Services**: A screenshot showing a "GCEM Test Jobs" configuration interface with a "Job Status" table.
- User Account Management and Customization**: A screenshot showing a "GCEM Grid Certificate Proxy" form with fields for "Certificate Pathname", "MyProxy Server", and "MyProxy Port".
- Security Services (SSH, Grid Security Infrastructure)**: A screenshot showing a "GCEM Grid Certificate Proxy" form with fields for "Certificate Pathname", "MyProxy Server", and "MyProxy Port".

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

The image shows two overlapping screenshots of the CyberWeb interface. The top screenshot is the home page, featuring a navigation bar with 'GCOM', 'MyGCOM', 'Execution', 'Data', 'Info & Status', and 'Admin'. Below the navigation bar, there are sections for 'News and Messages' with a timestamp of 'Apr 09,10 08:57 AM' and a message: 'You have successfully installed CyberWeb.' A 3D bar chart is visible on the right side. The bottom screenshot shows a 'Change your personal information' form with fields for 'username', 'lastname', 'institution', 'firstname', and 'email'. The form is partially filled with the values 'mary', 'thomas', 'SDSU', 'mary', and 'mary@gmail.com'. Below the form are buttons for 'Save User Information' and 'Cancel'. There are also sections for 'CyberWeb Preferences' and 'Login Statistics'.

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

The screenshot displays the GCEM Portal interface. At the top, there is a navigation bar with a host dropdown set to 'dolphin' and a directory field containing '/nfs/dolphinfs/home2/mthon'. Below this is a table with columns for 'Name', 'Size', and 'Modified'. The table lists a 'Parent Directory..' and several sub-directories numbered from 000001 to 000077. A second window is open, showing a similar interface for the host 'longboard.acel.sdsu.edu' and directory '/u08/cwproj/users/mary/201'. This window also has a table with columns for 'Name', 'Size', and 'Modified', listing a 'Parent Directory..' and a file named 'job.info.000168'. To the right of the second window is a 'Gcom Results Image' section containing a 3D visualization titled 'Vertical x-z Plane 15, Velocity Slide 7'. Below the image is a 'Visualization Scripts' section with fields for 'jobname', 'jobid', and 'jobstate', and a 'Run Visualization Script' button.

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

The screenshot displays the CyberWeb interface with several overlapping windows:

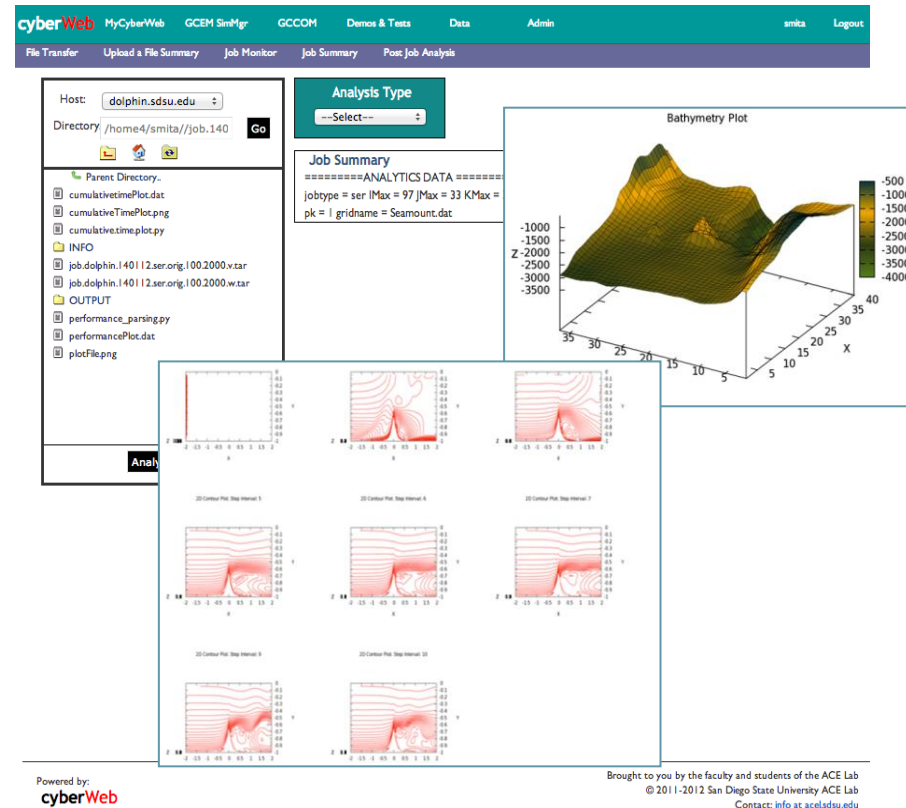
- Top Window (GCOM Test Jobs):** Shows a list of test jobs (Lid Drvn Cav. II, Temperature I, Temperature II, Seamount I) and a selection of a 'Lid Driven Cavity Test Case' with a 3D model.
- Middle Window (Job Configuration):** Shows the configuration for 'GCOM Test Runs: Temperature Test Case 1 For user: mary'. It includes fields for Job Name, Job Description, Select Host, and Job Details.
- Bottom Window (Messages):** A table showing system messages:

Date	Message
April 09,2010	You have successfully installed CyberWeb.
April 09,2010	Welcome to CyberWeb.
- Bottom Window (Queued Jobs):** A table showing the status of queued jobs:

ID	Name	Service	Resource	Status	Submit Time	Start Time	End Time
000001	ldc_demo_001	GSISSH	pipeline3	0	04/09/10 08:59:35		
000002	ldc_demo_001	GSISSH	pipeline3	0	04/09/10 10:00:15		
000003	ldc_demo_001	GSISSH	pipeline3	0	04/09/10 10:00:15		
000004	ldc_demo_001	GSISSH	pipeline3	0	04/09/10 10:00:15		
- Bottom Window (Running Jobs):** Shows 'No jobs running.'

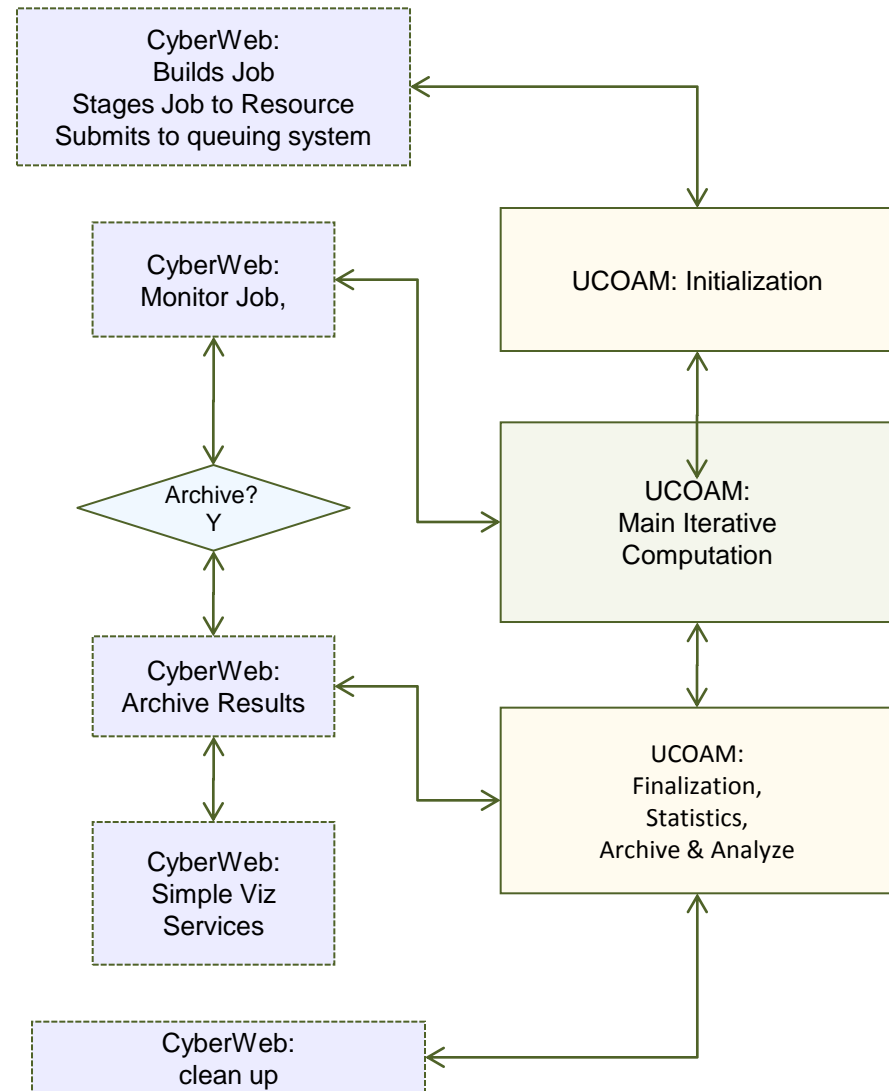
CyberWeb Visualization (CyberViz)

- 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



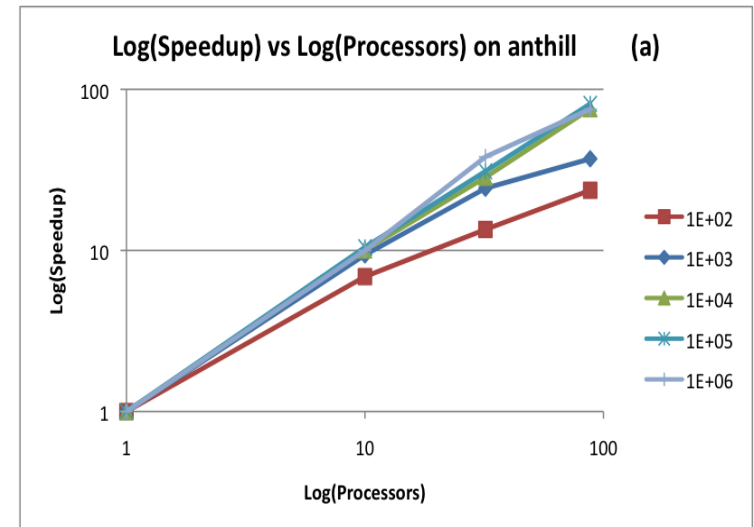
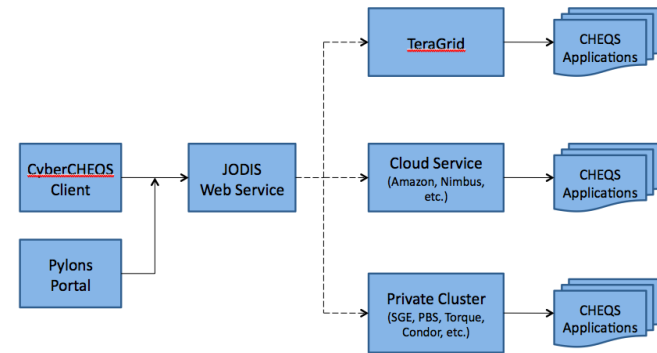
CyberWeb App: Running Parallel UCOAM Applications

- 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 reactive 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.