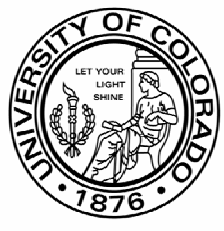
Unidata Seminar
Monday 22 January 2006

Grid-BGC: A Grid Enabled Carbon Cycle Modeling Environment

Jason Cope and Matthew Woitaszek
University of Colorado, Boulder
Jason.Cope@colorado.edu
Matthew.Woitaszek@colorado.edu



NCAR

Motivation: NCAR as an Integrator

- Scientific workflows are becoming too complicated for manual (or semi-manual) implementation.
- Not reasonable to expect a scientist to routinely:
  - Design simulation solutions by chaining together application software packages
  - Manage the data lifecycle (check out, analysis, publishing, and check in)
  - Do this in an evolving computational and information environment
- **NCAR must provide the software infrastructure to allow scientists to seamlessly (and painlessly) implement workflows**

Motivation: Robust Modeling Environments

- Our goal is to develop a simple, production quality modeling environment for NCAR and the geoscience community that insulates scientists from the technical details of the execution environment
  - Cyberinfrastructure
  - System and software integration
  - Data archiving

- Grid-BGC is an example of such an environment and is the first of these environments developed for NCAR
  - Learning as we develop and deploy
  - Tasked by the geoscience community, but developed services are applicable to other collaborative research projects

Outline

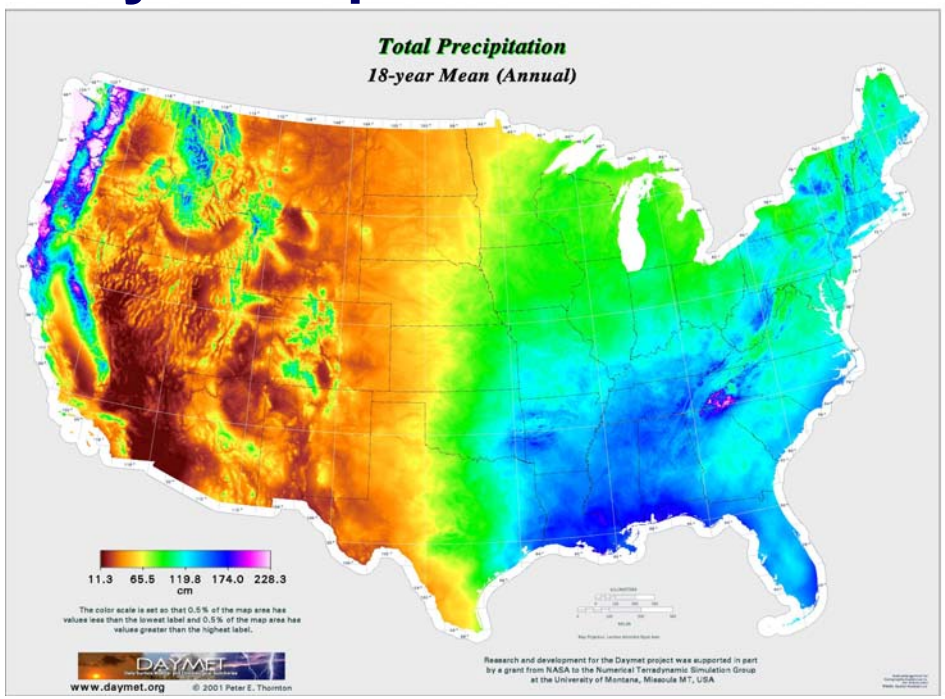
- Introduction
- Carbon Cycle Modeling
- Service Oriented Architecture for the Earth Sciences
- Grid-BGC System Architecture
- Re-tasking the services for other Earth Science applications
- Future Work

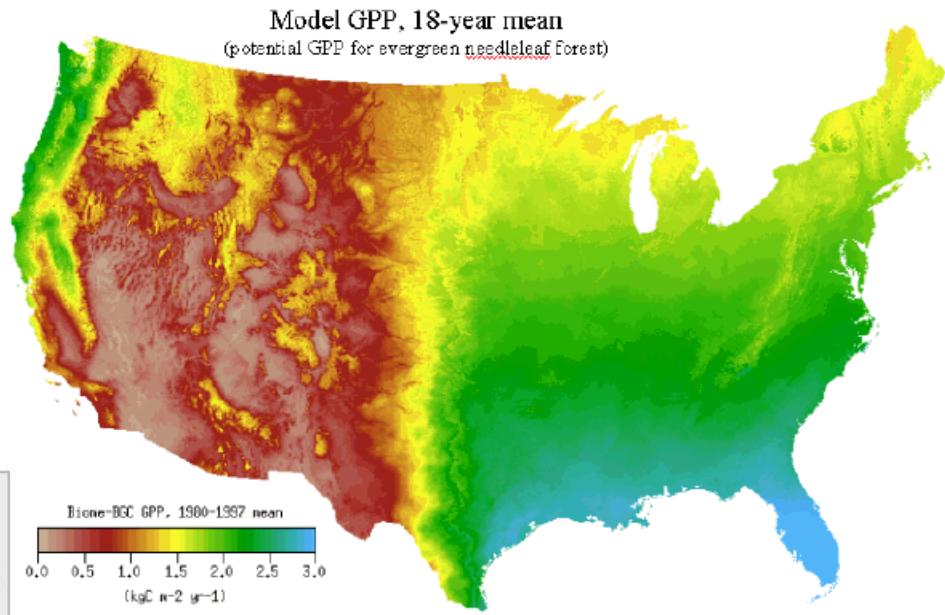
Introduction: Participants

- This is a collaborative project between the National Center for Atmospheric Research (NCAR) and the University of Colorado at Boulder (CU)
- NASA has provided funding for three years via the Advanced Information Systems Technology (AIST) program
- Researchers:
  - Peter Thornton (PI), NCAR
  - Henry Tufo (co-PI), CU
  - Luca Cinquini, NCAR
  - Jason Cope, CU
  - Craig Hartsough, NCAR
  - Rich Loft, NCAR
  - Sean McCreary, CU
  - Don Middleton, NCAR
  - Nate Wilhelmi, NCAR
  - Matthew Woitaszek, CU

Carbon Cycle Modeling: Workflow

Daymet inputs…



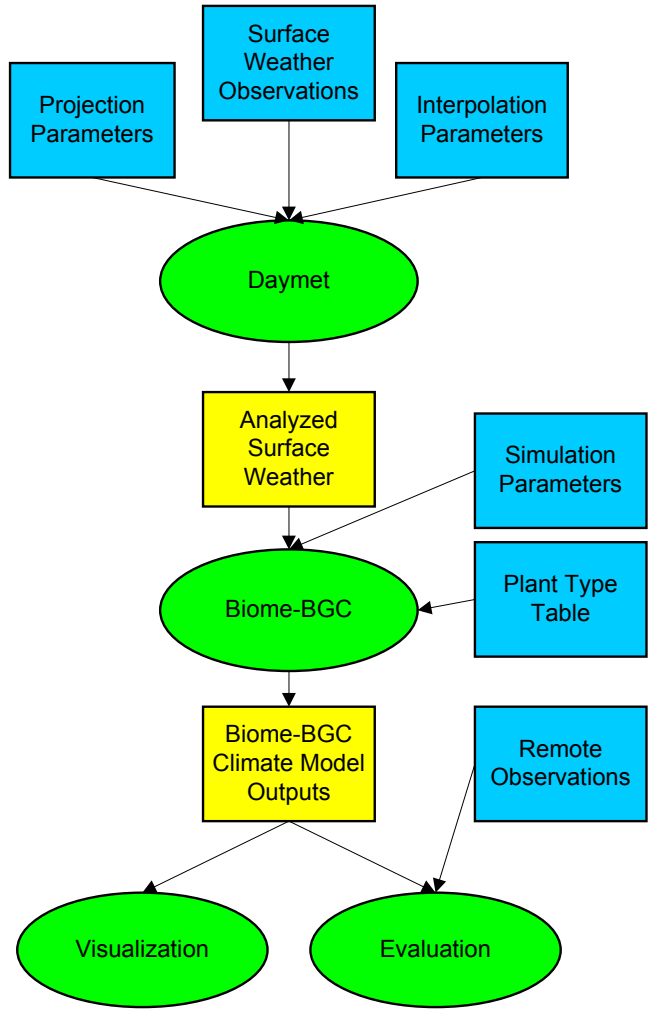


…Grid-BGC outputs

Carbon Cycle Modeling: Workflow

- **Daymet** model interpolates a high resolution grid of weather observations for a region
- **Biome BGC** model calculates carbon cycle parameters at each grid point

- Models originally intended for analysis of small geographic regions.
- Analysis of larger regions is accomplished by simulating its composite regions



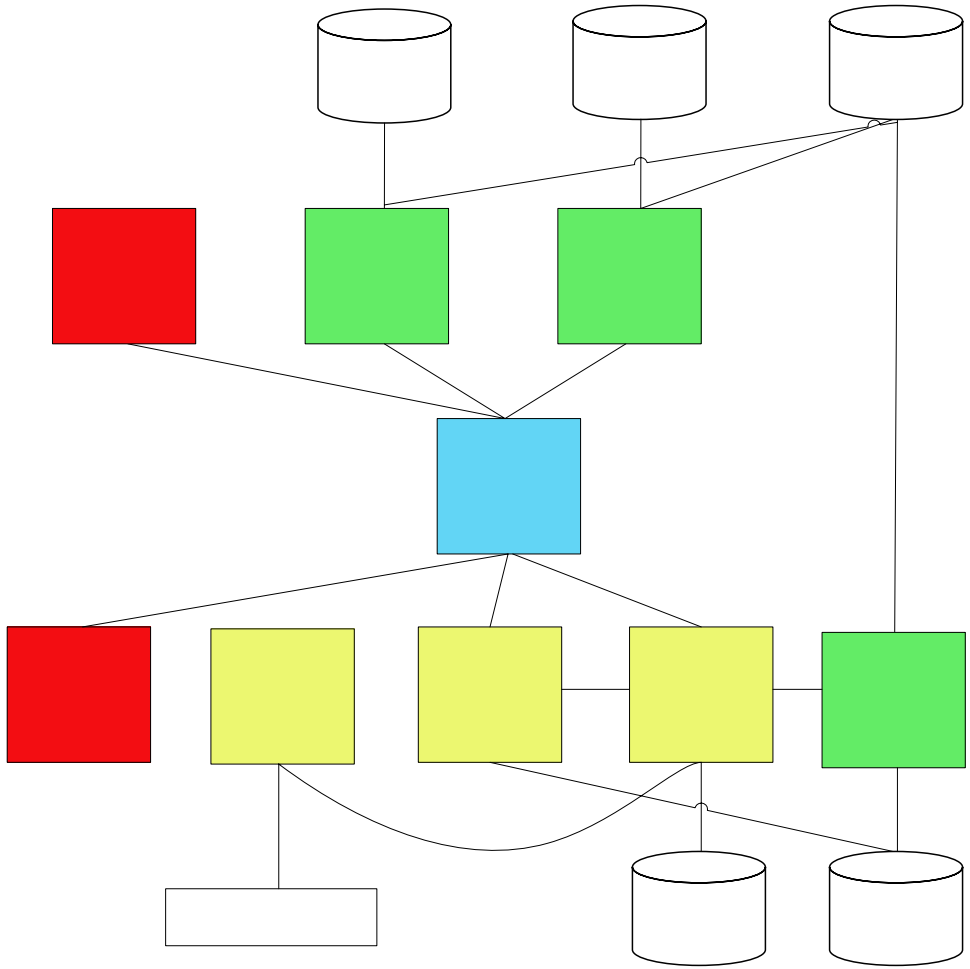
Carbon Cycle Modeling: Grid-BGC Motivation

Goal: Create an easy to use computational environment for scientists running large scale carbon cycle simulations.

- Requires managing multiple simultaneously executing workflows
  - Task creation
  - Execution management
  - Data management
- Distributed resource access across multiple organizations
  - Data archive and front-end portal are located at NCAR
  - Execution resources are located at CU and possibly other sites
- Reuse of software infrastructure
  - Extending the Grid-BGC workflow
  - Enabling other NCAR scientific applications and workflows

Service Oriented Architecture for the Earth Sciences: Desired Service Overview

- User interface services
  - Portal
  - GUI
  - Command line client
- Data services
  - Mass storage service
  - File transfer service
  - Data publishing service
- Execution services
  - Model execution service
  - Workflow control service
  - Resource allocation service
- Metadata services
  - Registry / Index Service
  - Resource brokerage service

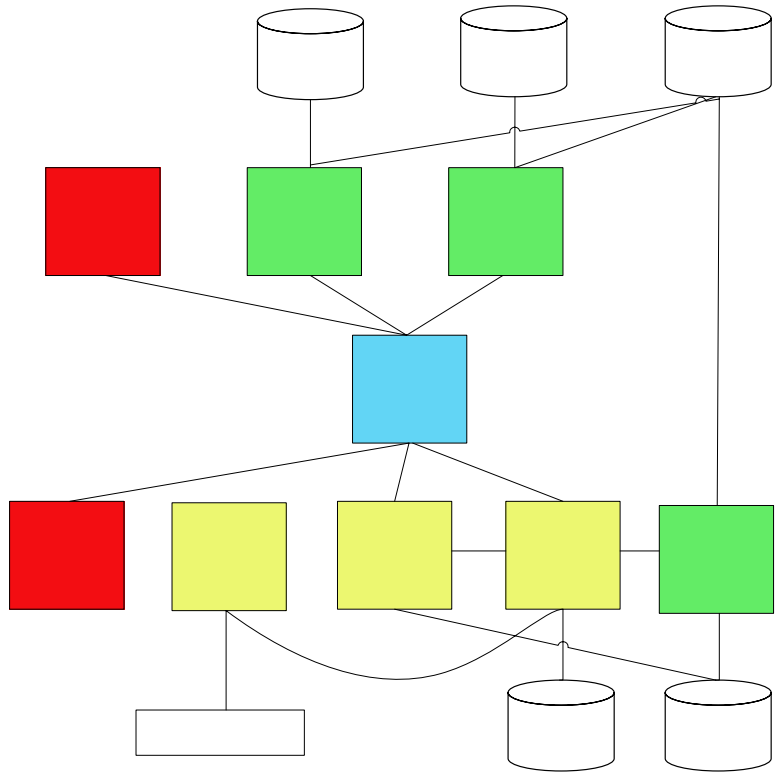


Grid-BGC: System Overview

- System goals
  - Easy to use
  - Efficient and productive science
- Development summary
  - Prototype developed with GT 3.2
  - Current system redeveloped with GT4
  - Integrates resources from NCAR and CU
- Architecture Implementation
  - Production system is not a pure service oriented architecture
  - Research and development system is a service oriented architecture

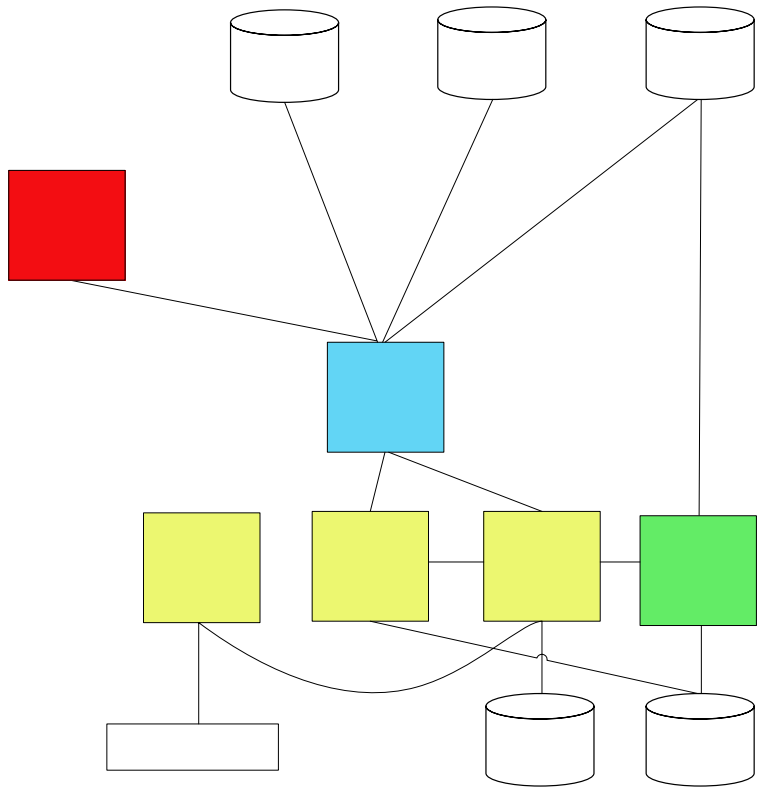
Service Oriented Architecture for the Earth Sciences: Implemented Services

- User interface services
  - Portal
  - GUI
  - Command line client
- Data services
  - Mass storage service
  - File transfer service
  - Data publishing service
- Execution services
  - Model execution service
  - Workflow control service
  - Resource allocation service
- Metadata services
  - Registry / Index Service
  - Resource brokerage service

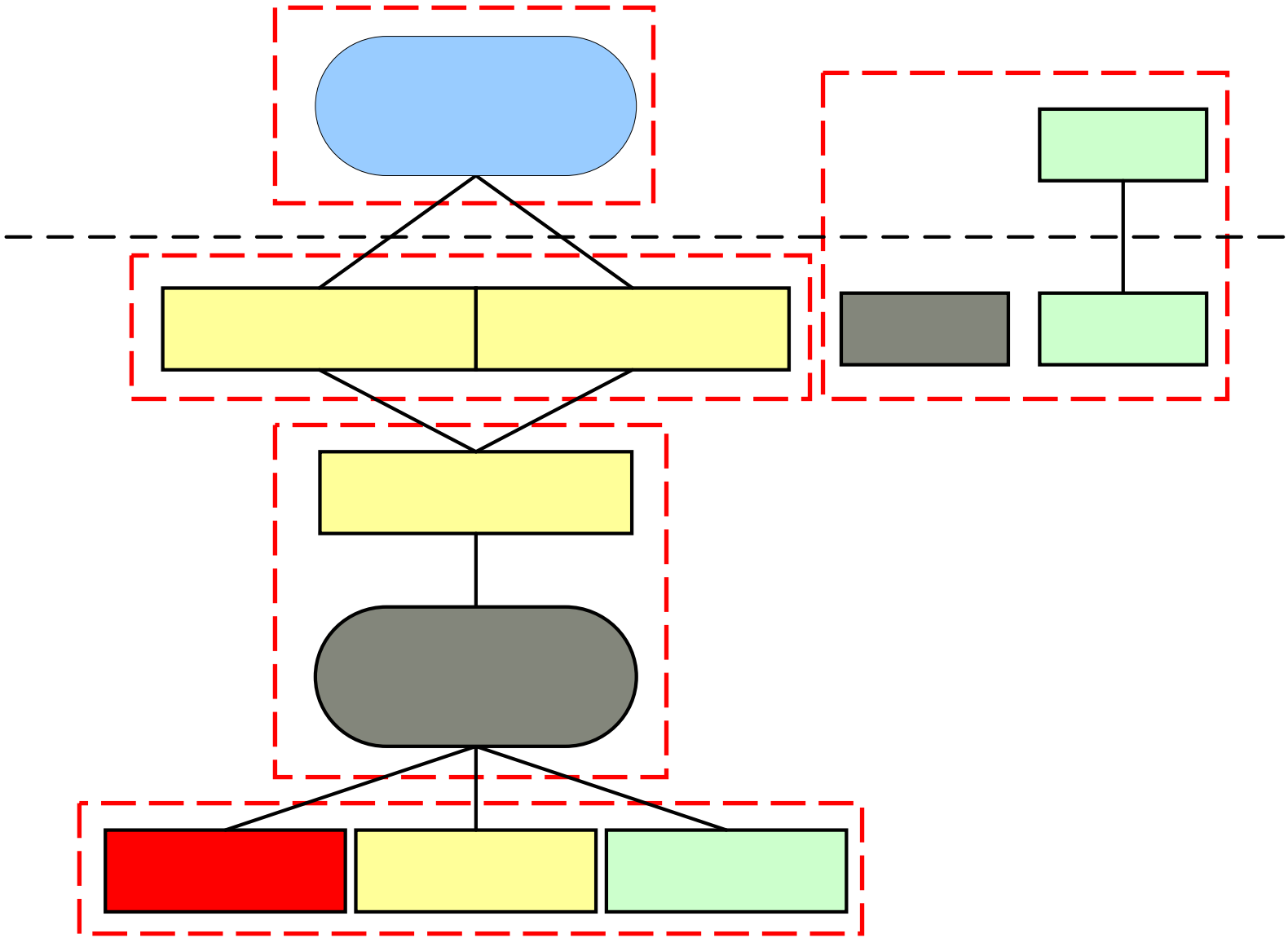


Service Oriented Architecture for the Earth Sciences: Implemented Services

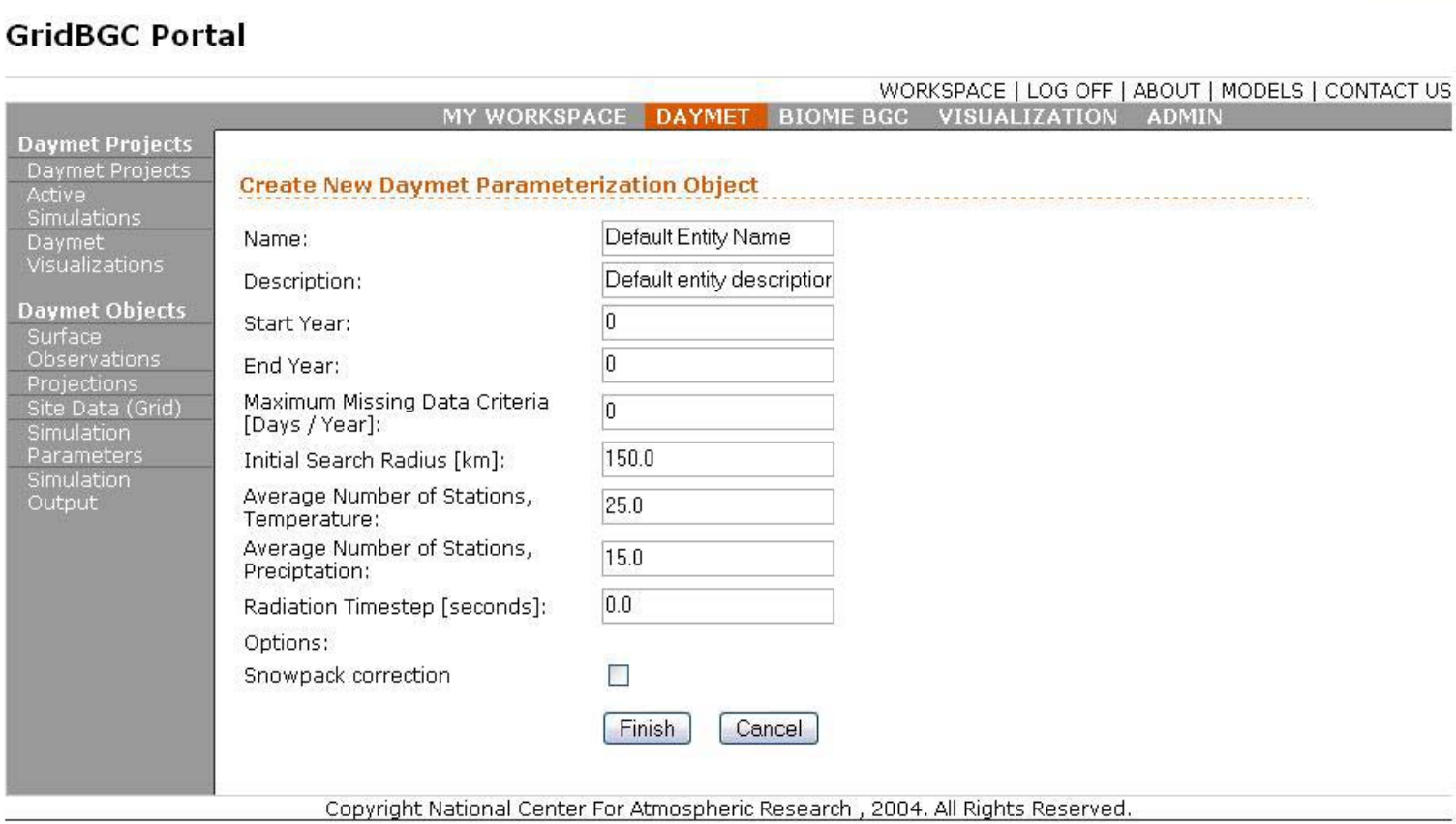
- User interface services
  - Portal
  - GUI
  - Command line client
- Data services
  - Mass storage service
  - File transfer service
  - Data publishing service
- Execution services
  - Model execution service
  - Workflow control service
  - Resource allocation service
- Metadata services
  - Registry / Index Service
  - Resource brokerage service



Grid-BGC: System Architecture



Grid-BGC Portal



- Web interface to Grid-BGC
- JSP / Tomcat implementation using CoG Kit
- Composed of logical services

Grid-BGC Execution Services

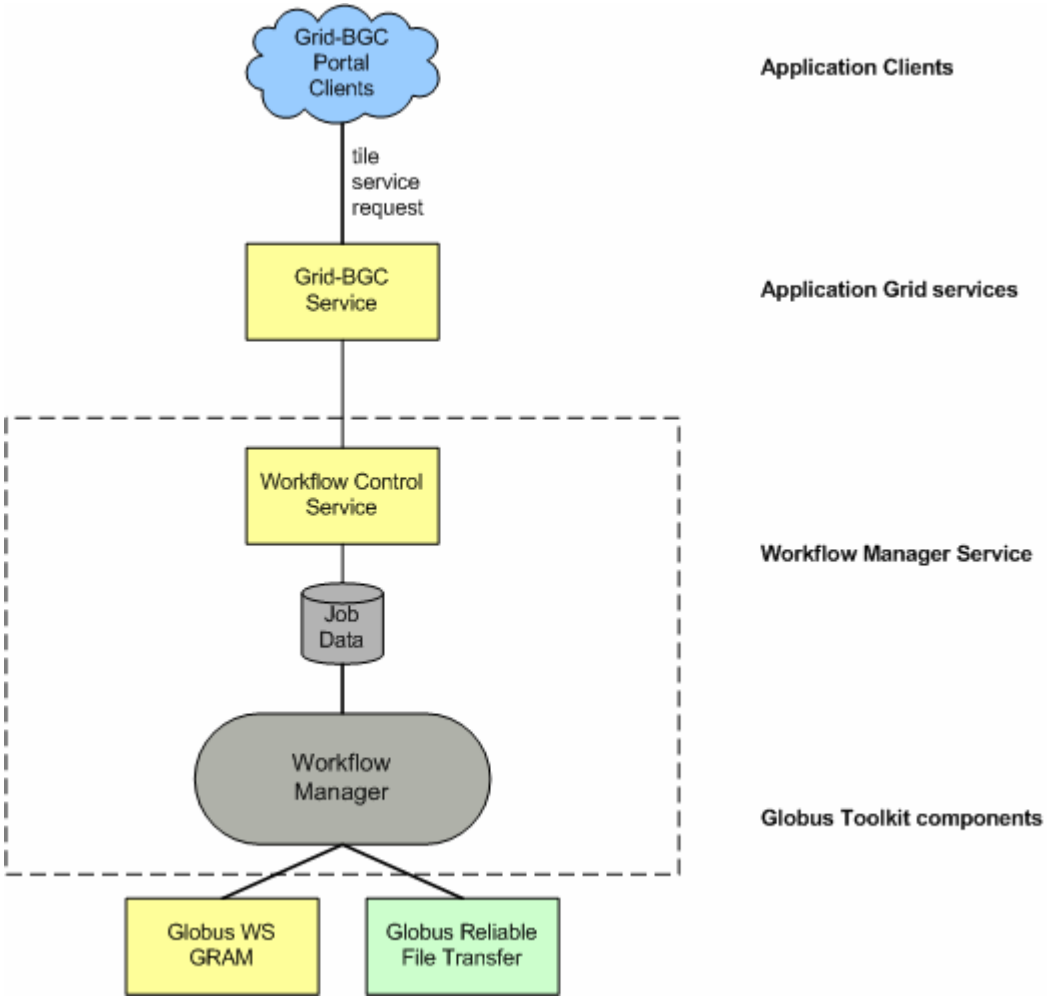
- Execution service contains all functionality needed to run a model and is aware only of those models
- Provides interface to request and initialize a model run
  - Creates directory structure
  - Creates model initialization files
  - Registers file transfers and executables with the workflow manager
- Provides interfaces to query, terminate, and cleanup requested model runs

Workflow Control Service and Workflow Manager

- Workflow Control Service
  - Provides functions to register workflow tasks, model executions, and file transfers
  - Execution service uses the workflow control service functions to register its tasks
  - Workflow control service stores the workflow metadata in a persistent database

- Workflow Manager
  - Periodically queries the workflow metadata database for new tasks to execute
  - Delegates file transfers to the Reliable File Transfer service (RFT) and job executions to the Grid Resource and Allocation Management Service (GRAM)

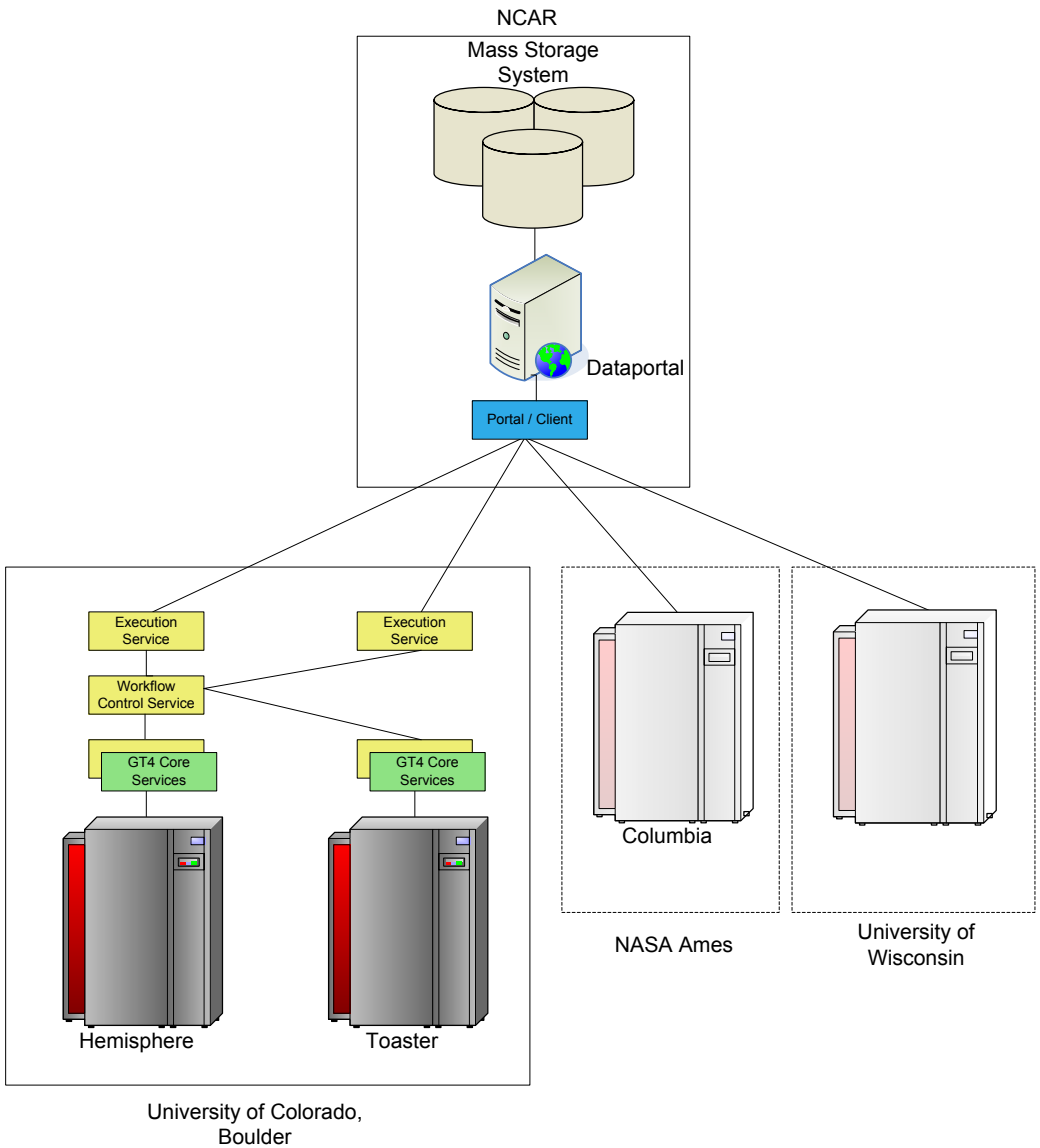
Example Grid-BGC Workflow



Operational Experience

- User Interface has been externally beta tested
  - Beta testers from
    - University of Wisconsin
    - Utah State University
    - WSL Switzerland
  - Feedback helped improve users interactions with the portal
- Grid computing and modeling environment beta tested internally
  - Short term productivity gains have been realized using this system

Current Grid Topology



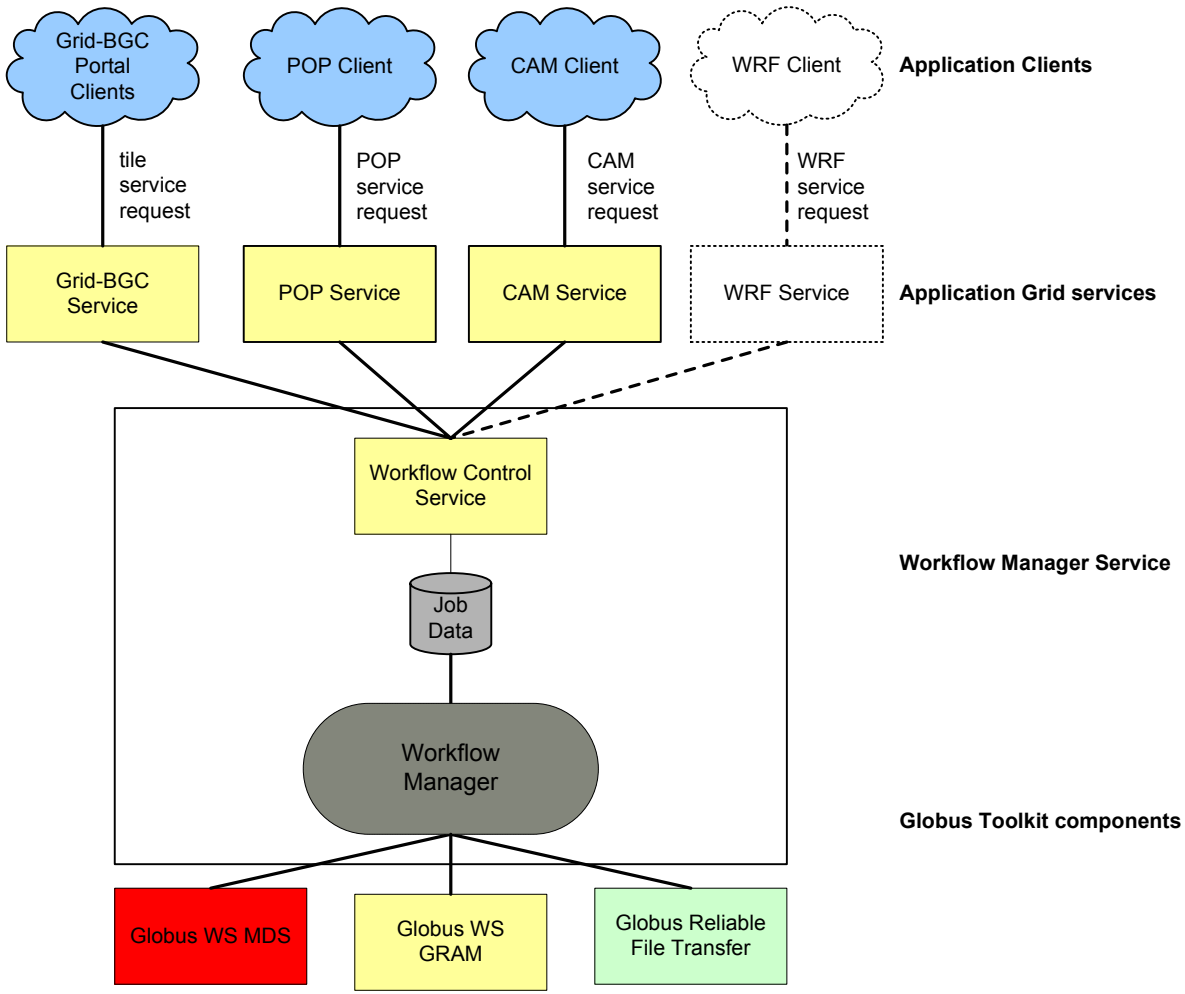
Grid Enabling CAM and POP

- Community Atmosphere Model (CAM)
  - Developed by NCAR
  - Atmospheric component of NCAR's Community Climate System Model (CCSM)
- Parallel Ocean Program (POP)
  - Developed by the DOE at the Los Alamos National Laboratory
  - Ocean component of CCSM
- Grid Enabling CAM and POP
  - Re-tasked the grid service and workflow subsystem to run CAM and POP
  - New components
    - Execution services
    - Client interfaces for accessing the services
  - Reused components
    - Workflow subsystem and service
    - Service registry
    - Service communication package

Future Work: Expansion of the Grid-BGC Environment

- Integrate new computational resources
  - Integrate NASA's Columbia Supercomputer into the Grid-BGC environment
  - Integrate resources provided by the system's users (University of Wisconsin, …)
  - TeraGrid
- Continue to break out the desired services from current system components
- Continue to evolve system architecture into a service oriented architecture (SOA)
- Visualization

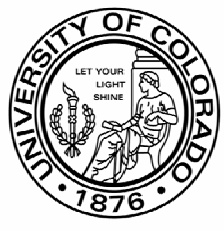
Future Work: Grid Enabling More Earth Science Applications



Grid-BGC: A Grid Enabled Carbon Cycle Modeling Environment

This research was supported in part by the National Aeronautics and Space Administration (NASA) under AIST Grant AIST-02-0036, the National Science Foundation (NSF) under ARI Grant #CDA-9601817, and NSF sponsorship of the National Center for Atmospheric Research.

Questions?
Ideas? Comments?
Suggestions?
http://www.gridbgc.ucar.edu



Presenter's email:
Jason.Cope@colorado.edu

NCAR