

# Working Meeting at FermiLab

University of California Grid (UC Grid)

January 19-20, 2010

Fermi Lab, Chicago, IL

**LBNL**

Krishna Muriki, Jackie Scoggins, Tammy Welcome,  
Shreyas Chola, Doug Olson, Gary Jung, Greg Kurtzer  
Lawrence Berkeley National Laboratory



William Strossman, Russ Harvey, Bob Grant  
University of California, Riverside



Harry Mangalam, Joseph Farran, Frank Wessel  
University of California, Irvine



Kejian Jin, Shao-Ching Huang, TV Singh, Sveta Mazurkova, Bill  
Labate, Qiyang Hu, Prakashan Korambath, Seonah Kim, Scott  
Johnson, Puneet Souda



Paul Weakliem, Arlene Allen  
University of California, Santa Barbara



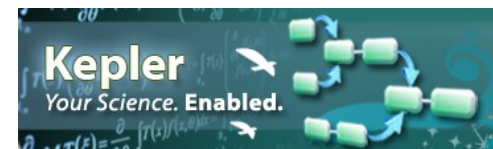
David Walker, Bill Broadley, Russ Hobby  
University of California, Davis



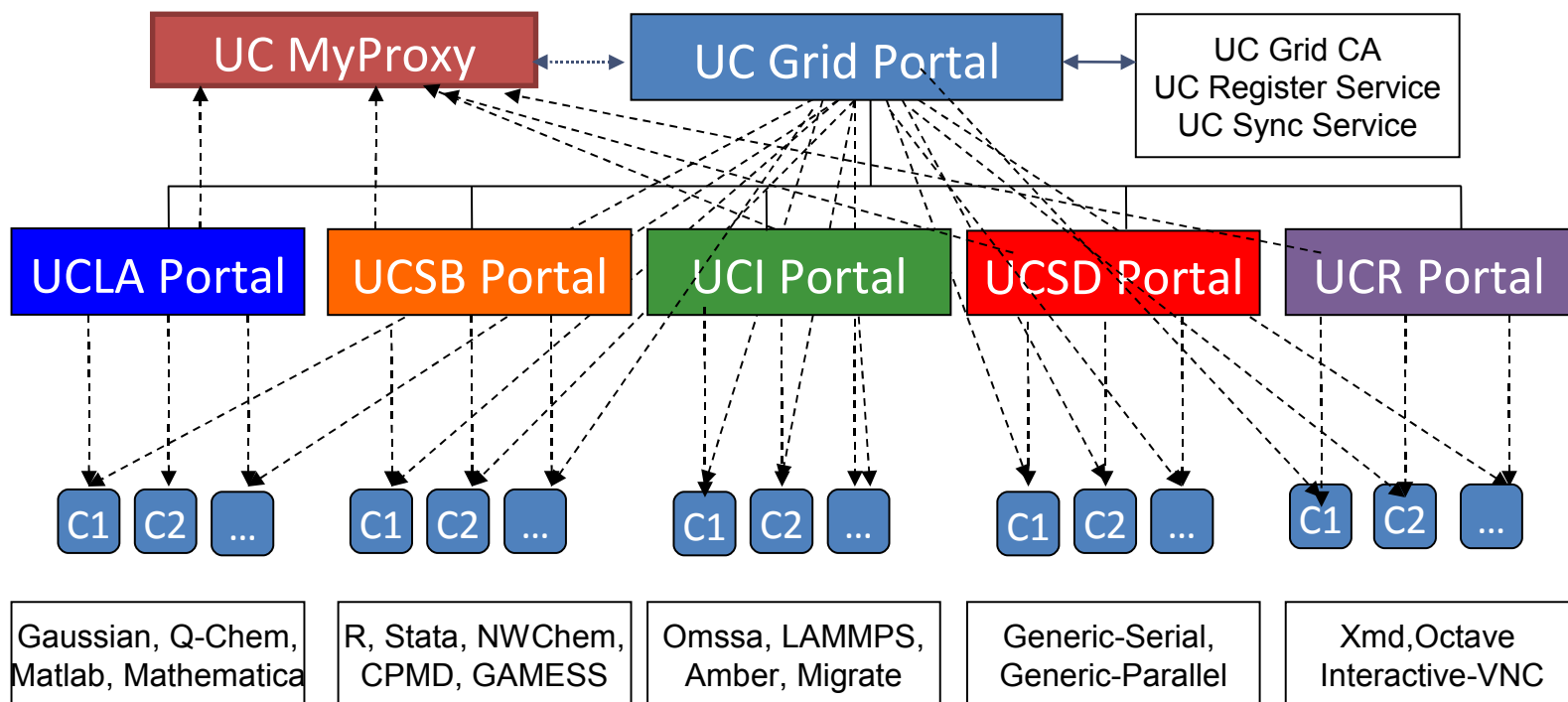
Shava Smallen, Jianwu Wang, Ilkay Altinas,  
Richard Moore, David Hutches  
San Diego Super Computing Center



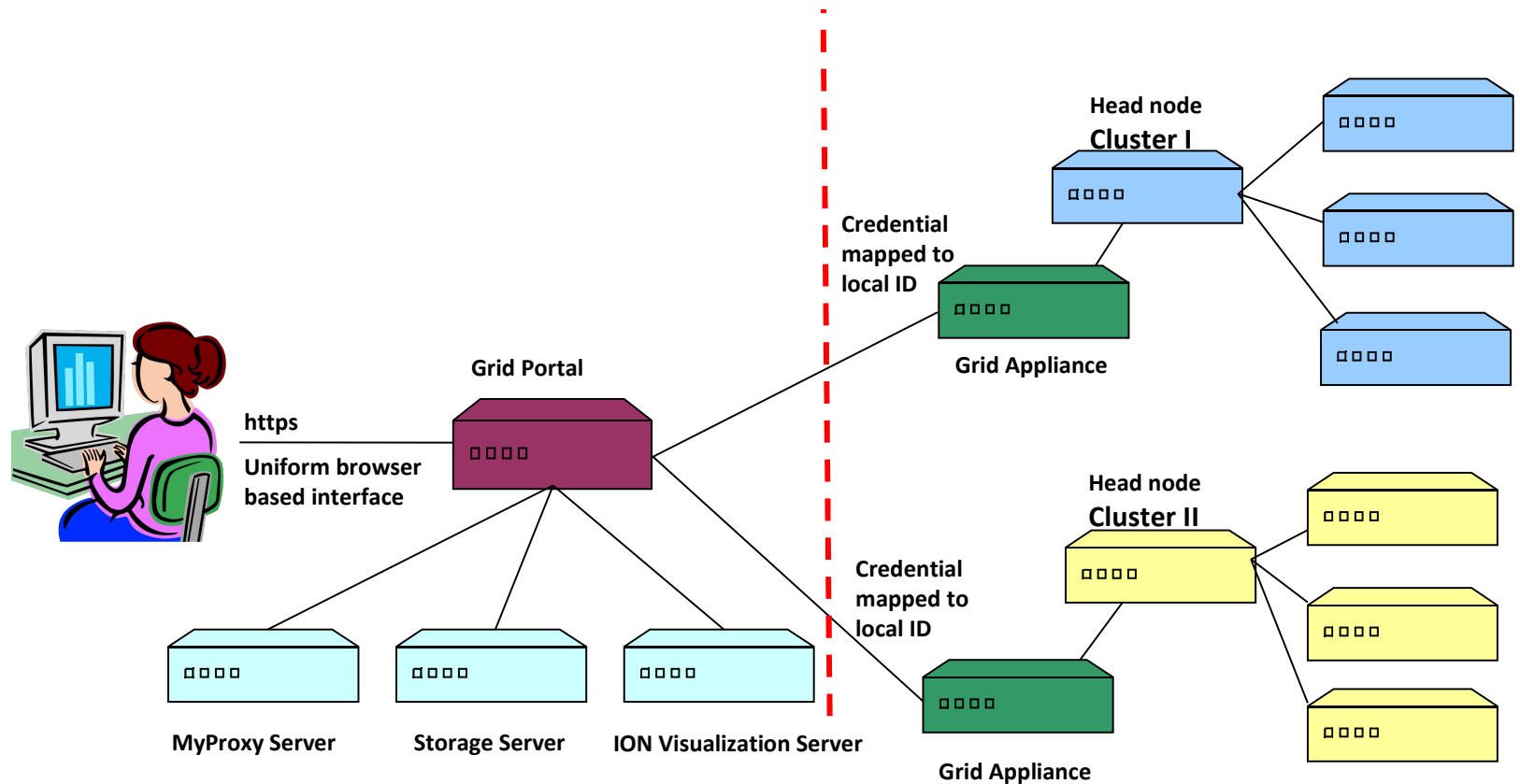
**SDSC**



# UC Grid Architecture



# Single Campus Architecture



# The UC Grid Portal

- Provides Register Service (SOA):
  - user portal account creation and uniqueness of user name in the UC Grid, automatic CA sign, MyProxy push.
- Provides Sync Service (SOA) between campus portal and UC Grid portal:
  - Any change of an user account at a campus Grid Portal is automatically propagated to the UC Grid Portal.
  - Any change of new resources at a campus Grid Portal will also be reflected at UC Grid Portal
- From the UC Grid Portal, users can:
  - work with clusters from all the campuses those users are authorized to use.
  - transfer files across the campuses.
  - Use pooled resources across the campuses.

# Campus Grid Portal

- Features:
  - Resource Discovery
    - Cluster load, usage, queue information, job details
  - Data Manager
    - File management such as create, remove, rename, edit, transfer etc.
  - Job Services
    - Submit job as a cluster user
    - Submit job as a pool user
  - Interactive X window for applications
    - Matlab, Mathematica, Maple, Xstata etc.
  - Access to Teragrid

# Job Submission

- General Job submission
  - Any user created executable
- Application job submission
  - Predefined executable
    - As Cluster user
    - As Pool user
- Pool user application job submission
  - UGP Chooses the target cluster for the application job.
  - Stages the input files to a guest user login id at the target cluster
  - Submits the job to run under a guest login id at the target cluster.
  - Local scheduler determines when those jobs are run.
  - UGP facilitates the transfer of the output files back to the Pool User.
  - Provide job monitoring and auditing

# Interactive Xterm through VNC

The screenshot displays a web-based VNC interface. At the top, a navigation bar includes tabs for 'Welcome', 'Administration', 'Grid Admin', 'My Home', 'Resources', 'Data Manager', 'Job Services', 'Other Grids', and 'Interactive'. Below this, a header bar shows 'Interactive Apps', 'Development Environment', and 'SSI'. The main content area is divided into two panels. The left panel, titled 'Interactive Applications', lists various clusters and their associated applications: Neutrino Cluster (Xterm), Houk Cluster (Xterm), CCPR Cluster (Xterm), CNSI Cluster (Xterm), Hydro Cluster (Xterm), Springfield Cluster (Xterm), Cardio Cluster (Xterm), Hoffman2 Cluster (Xstata VNC, Xterm, OpenDX VNC, Mathematica VNC, Xmaple VNC, GnuPlot VNC, Molden VNC, Xmgrace VNC, Matlab VNC, TecPlot VNC, Abaqus VNC). The right panel shows a terminal window titled 'ppk@i01:~' with a list of files and directories, including 'iberrorhosts', 'ifcfg-eth0-home', 'ifcfg-eth1-home', 'IJ\_Prefs.txt', 'Imat.dat', 'ims1test', 'Inca', 'inca-reporters', 'in.na', 'in.water', 'iptables-home', 'irina', 'isoart', 'java.log.10699', 'java.log.31903', 'JOB-ARRAY', 'job.out', 'JobSubmission.xml', 'k42Data', 'kepler', 'Kepler-1.0.0', 'kepler-linux-1.0.0.jar', 'kepler-test', 'Keys', 'kishan', 'lapack++', 'LAPACK', 'lapack-3.1.1', 'lapack-c-routines', 'lapackpp-2.5.2', 'lapackpp-2.5.2.tar.gz', 'lapack.tgz', 'lapack++.tgz', 'lars', 'LES', 'lib', 'libtest', 'libX11.a', 'list', 'i01:~ {1002}\$'. The bottom status bar shows 'Application Name', 'Appliance Name', 'Person Name', 'Appliance Name', 'Date/Time', and 'User Name'.

Interactive Applications

To run the interactive application, your browser m

**Neutrino Cluster**  
Xterm

**Houk Cluster**  
Xterm

**CCPR Cluster**  
Xterm

**CNSI Cluster**  
Xterm

**Hydro Cluster**  
Xterm

**Springfield Cluster**  
Xterm

**Cardio Cluster**  
Xterm

**Hoffman2 Cluster**  
Xstata VNC  
Xterm  
OpenDX VNC  
Mathematica VNC  
Xmaple VNC  
GnuPlot VNC  
Molden VNC  
Xmgrace VNC  
Matlab VNC  
TecPlot VNC  
Abaqus VNC

ppk@i01:~

iberrorhosts  
ifcfg-eth0-home  
ifcfg-eth1-home  
IJ\_Prefs.txt  
Imat.dat  
ims1test  
Inca  
inca-reporters  
in.na  
in.water  
iptables-home  
irina  
isoart  
java.log.10699  
java.log.31903  
JOB-ARRAY  
job.out  
JobSubmission.xml  
k42Data  
kepler  
Kepler-1.0.0  
kepler-linux-1.0.0.jar  
kepler-test  
Keys  
kishan  
lapack++  
LAPACK  
lapack-3.1.1  
lapack-c-routines  
lapackpp-2.5.2  
lapackpp-2.5.2.tar.gz  
lapack.tgz  
lapack++.tgz  
lars  
LES  
lib  
libtest  
libX11.a  
list  
i01:~ {1002}\$

test\_log.o132670  
test\_log.o132691  
testmfold  
test\_OMSSA  
test.pl  
test-suite  
THREAD  
tmp  
touched\_file  
touched\_it  
TSM  
tutorial  
Tutorial  
tvexamples  
twodips.out  
twoeigs.out  
twoeigvecs.out  
two\_r12.out  
typescript  
uccsc08\_sb2.ppt  
uccsc08\_sbns.ppt  
uccsc08\_sb.pptx  
uccsc08\_sbsh.ppt  
uccsc08\_ucgrid.pdf  
Usersupport  
utils  
veryshort.q  
vms  
vms.c  
wglover  
wisdom  
WORK3  
xfig.3.2.5.full.tar.gz  
XFree86  
xmdrun  
xmd-test  
xv-3.10a.tar.gz  
yugu

Application Name	Appliance Name	Person Name	Appliance Name	Date/Time	User Name
Xterm	grid.hoffman2.idre.ucla.edu	51002	ppk	Mon Apr 27 20:11:17 PDT 2009	



# Interactive GUI Application through VNC

**UCLA Grid Portal**

My Home Resources Data Manager Job Services Other Grids **Interactive**

Interactive Apps Development

Interactive Applications

**Interactive Applications**

To run the interactive application, select a cluster and application from the list below.

**Hoffman Cluster**  
[Xterm](#)  
[Matlab](#)

**CNSI Cluster**  
[Xterm](#)

**Hydro Cluster**  
[Xterm](#)

**Dawson Cluster**  
[Xterm](#)

**Hoffman2 Cluster**  
[Xterm](#)  
[OpenDX](#)  
[Mathematica](#)  
[Xmaple](#)  
[GnuPlot](#)  
[Molden](#)  
[Xmgrace](#)  
[Matlab](#)  
[TecPlot](#)  
[Abaqus](#)

**MATLAB 7.5.0 (R2007b)**

File Edit Debug Distributed Desktop Window Help

Current Directory: /u/home2/

Shortcuts How to Add What's New

Current Directory Workspace

All Files	Type	Size
.config	Folder	
.fontconfig	Folder	
.gconf	Folder	
.gconfd	Folder	
.globus	Folder	
.gnome2	Folder	
.gnome2_private	Folder	
.jmol	Folder	
.maple	Folder	
.Mathematica	Folder	

Command Window

New to MATLAB? Watch t

>> |

Applet VncViewer started grid.ucla.edu:9443

Application Name	Appliance Name	Portal Port	User Name	Start Time	Kill	Reconnect
Matlab	grid.hoffman2.idre.ucla.edu	51012	joan	Wed Feb 27 11:03:52 PST 2008		

# File Transfer

Grid Data Manager - Neutrino Cluster

Connect File Edit View Help

/u/home4/ppk

ppk

Name	Date Modified	Size	Permissions
1	2009 02-27 1	6	-rw-r--r--
a.out	2008 02-01 1	608822	-rwxr-xr-x
aimk	2005 03-11 1	68825	-rwxr-xr-x
ajax	2006 08-08 1	4096	-rwxr-xr-x
alex.c	2004 12-20 1	129	-rw-r--r--
aphost-template	2006 06-05 1	760	-rw-r--r--
atssync	2006 03-07 1	4096	-rwx

Grid FTP Client

File Help

Hoffman2 Cluster Neutrino Cluster

Name	Size	Permissions
08Dec12-16-33-44_1.ct	2770	-rw-r--r--
08Dec12-16-33-44_1.pc	7964	-rw-r--r--
08Dec12-16-33-44_1.ps	1579	-rw-r--r--
10-11	2048	-rwxr-xr-x
11-12	2048	-rwxr-xr-x
110.cmd	1899	-rwxr--r--
12-3	2048	-rwxr-xr-x

Name	Size	Permissions
1	6	-rw-r--r--
a.out	6088	-rwxr-xr-x
aimk	6882	-rwxr-xr-x
ajax	4096	-rwxr-xr-x
alex.c	129	-rw-r--r--
aphost-template	760	-rw-r--r--
atssync	4096	-rwx

This is a debug window

# Generic Batch Job Submission Interface

[Welcome](#) | [Administration](#) | [Grid Admin](#) | [My Home](#) | [Resources](#) | [Data Manager](#) | **[Job Services](#)** | [Other Grids](#) | [Interactive](#)

**Job Status** | **Generic Jobs** | **Applications** | **Multi-Jobs** | **User Applications**


**Job Services**

**Generic Job Submission**

Submit to: Hoffman2 Cluster

**Job To Submit**  
Required entries have **bold** labels.


Job Name:  Just a name you give this job so you can recognize it later.

**Executable:**   The file name of your executable required.

Arguments:

Directory

- Directory is the directory in which your job will run. Every filename used in your job which is not specified as an absolute path will be relative to this directory.
- If Directory is omitted, your home directory will be used by default.
- If Directory is specified but does not include an absolute path, it will be relative to your home directory.

Directory:  

**JobType:** Serial

Environment Variables:

Stdin: <

**Job Requirements**

- For serial jobs, the number of processors must be one.
- Some schedulers terminate jobs that have reached their maximum CPU time, others use the maximum elapsed time.

Number of Processors:

Maximum Memory(MB):

Maximum Time (in hours):

Queue Name (Optional):

**Save as Your Application**  
Are you submitting the same job over and over again? Tired of filling in this form? Save your filled in form as an application. Then you can select and submit it from "User Applications".

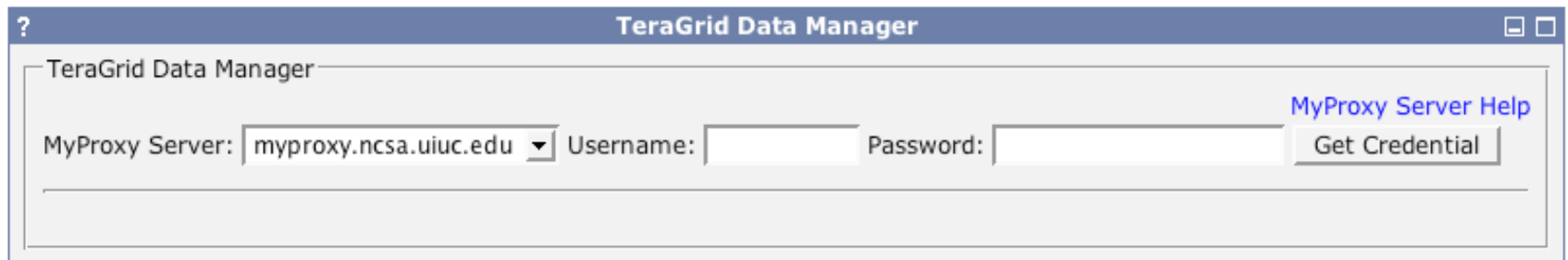
Submit Job | ☐ Save as your application!

# Predefined Application

Welcome Administration Grid Admin My Home Resources Data Manager <b>Job Services</b> Other Grids Interactive	
<b>Job Status Generic Jobs Applications Multi-Jobs User Applications</b>	
<b>Applications</b>	
Application Submission	
Predefined Applications	Amber10 parallel
<b>Neutrino Cluster</b> Mathematica  Mathematica 64 bit	Submit to: Hoffman2 Cluster
<b>Houk Cluster</b> Gaussian03-Parallel  Gaussian03-Serial	<b>Job To Submit</b> Required entries have <b>bold</b> labels.
<b>CCPR Cluster</b> Stata	Application Description: A molecular dynamics software package
<b>CNSI Cluster</b> Gaussian03-Serial  Q-Chem	Job Name: <input type="text" value="Amber10-Hoffman2-parallel"/> Just a name you give this job so you can recognize it later.
<b>Hoffman2 Cluster</b> Gaussian  Amber10 parallel  Amber10 serial  Q-Chem-Parallel  R  Amber9 parallel  Stata10  Amber9 serial	Arguments: <input type="text"/>  Directory <ul style="list-style-type: none"><li>• If you specify a directory, your job will be run in that directory.</li><li>• If you don't specify a directory, your job will run in your home directory.</li><li>• Unless an absolute path is specified for any file used in the job, the filename will be relative to the directory specified or your home directory, if omitted.</li><li>• Click here for more information</li></ul>
	Directory: <input type="text"/> Stdin: < <input type="text" value="/dev/null"/> <b>JobType:</b> <input type="button" value="MPI Parallel"/>
	<b>Job Requirements</b> <ul style="list-style-type: none"><li>• For serial jobs, the number of processors must be one.</li><li>• Some schedulers terminate jobs that have reached their maximum CPU time, others use the maximum elapsed time.</li></ul>
	Number of Processors: <input type="text" value="1"/> Memory Per Processor: <input type="text" value="1024"/> Maximum Time (in hours): <input type="text" value="1"/> Queue Name (Optional): <input type="text"/> Project Name(Optional): <input type="text"/>
	<b>Save as Your Application</b> Are you submitting the same job over and over again? Tired of filling in this form? Save your filled in form as an application. Then you can select and submit it from "User Applications".
	<input type="button" value="Submit Job"/> <input type="checkbox"/> Save as your application!

# Job Submission to Other Grid Resources

User interface to fill in username and password to retrieve short lived credential for submitting jobs to Teragrid

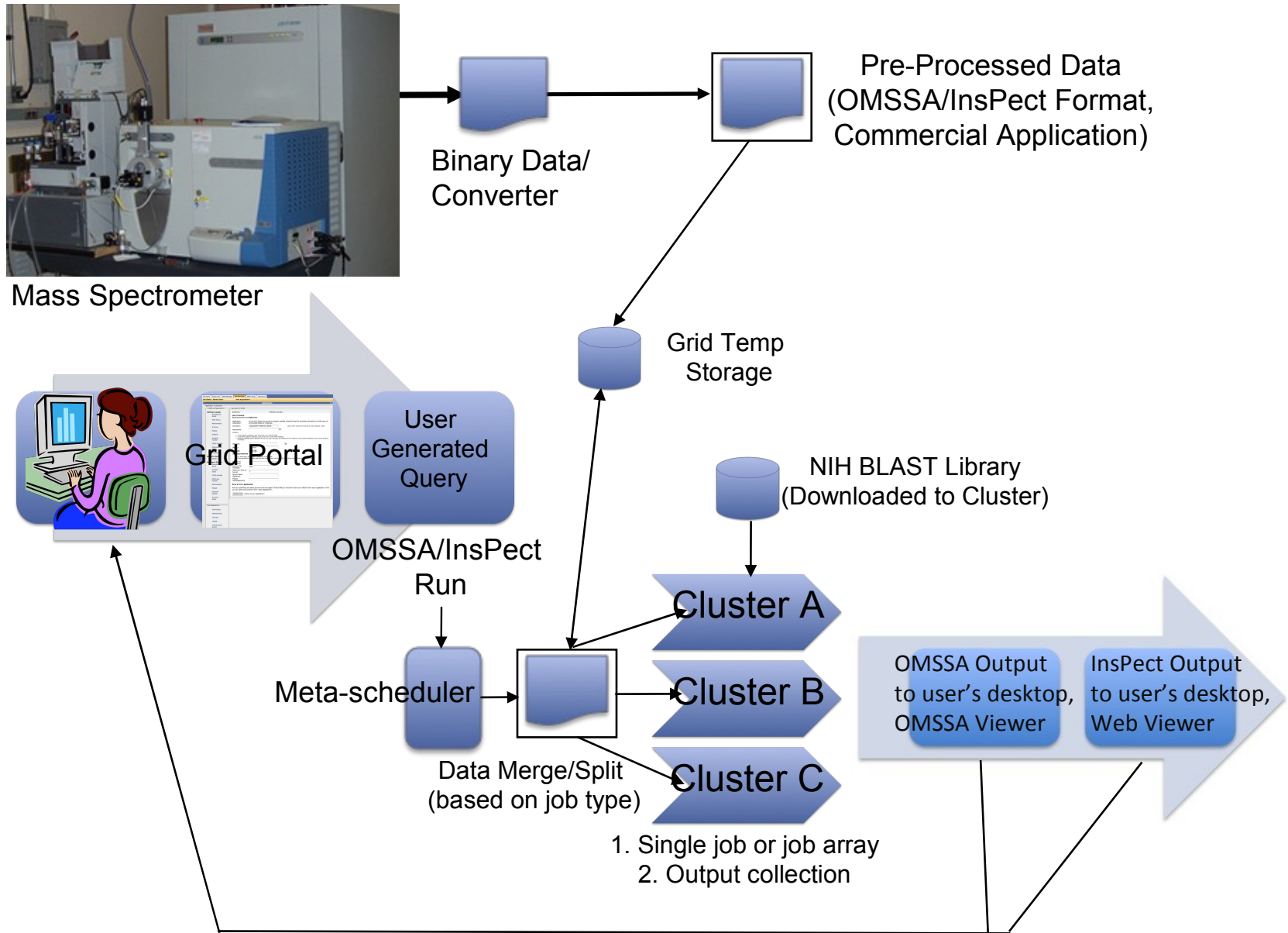


The screenshot shows a web browser window titled "TeraGrid Data Manager". The page has a header bar with a question mark icon on the left and window control icons on the right. Below the header, the text "TeraGrid Data Manager" is displayed. The main content area contains a form with the following elements:

- A label "MyProxy Server:" followed by a dropdown menu showing "myproxy.ncsa.uiuc.edu".
- A label "Username:" followed by a text input field.
- A label "Password:" followed by a text input field.
- A blue hyperlink "MyProxy Server Help" located to the right of the password field.
- A button labeled "Get Credential" positioned to the right of the password field.

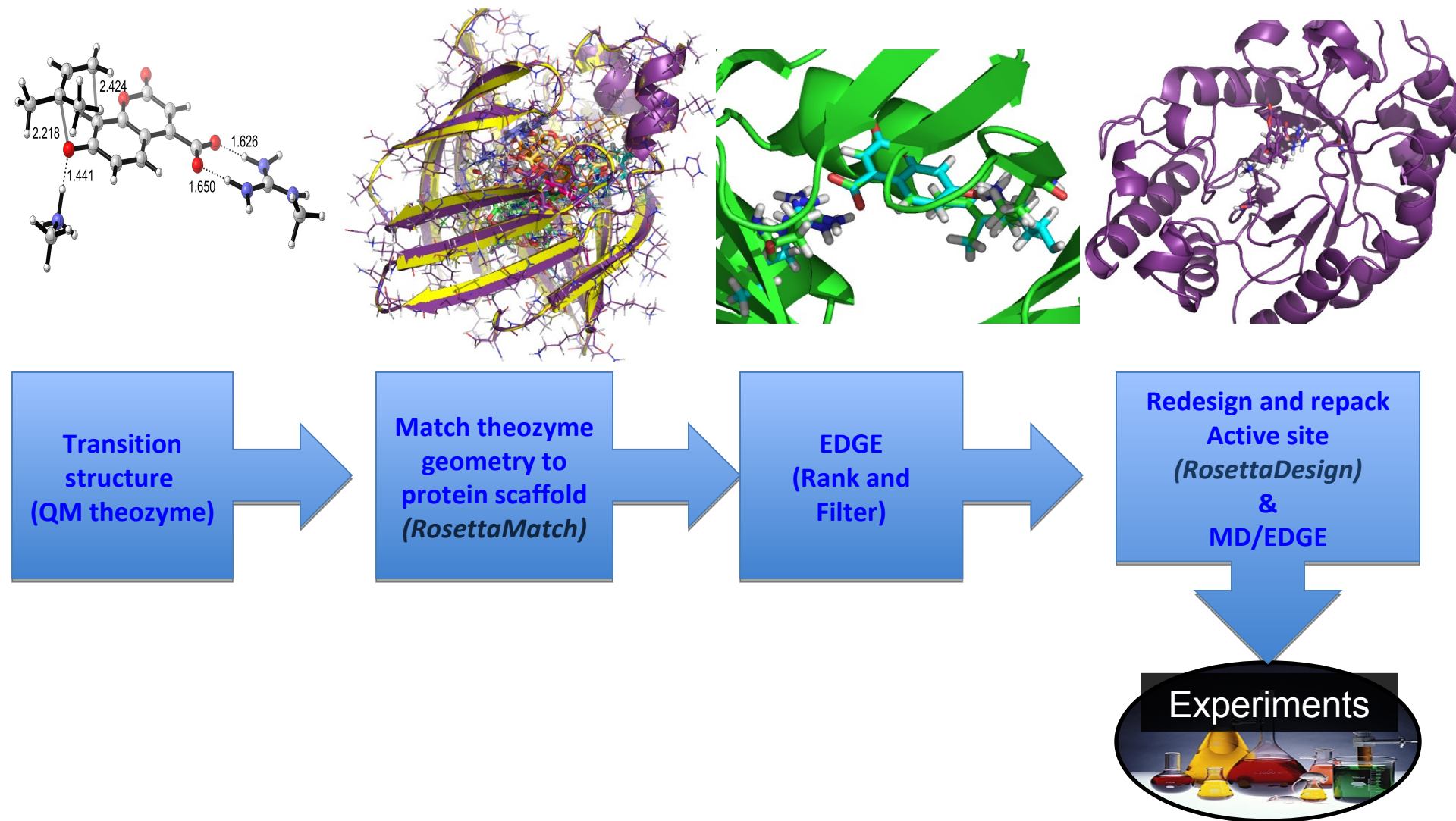
Below the form fields, there is a horizontal line.

# Proteomics Workflow Project



## Inside-out protocol to design novel enzyme catalysts

Baker - Houk collaboration

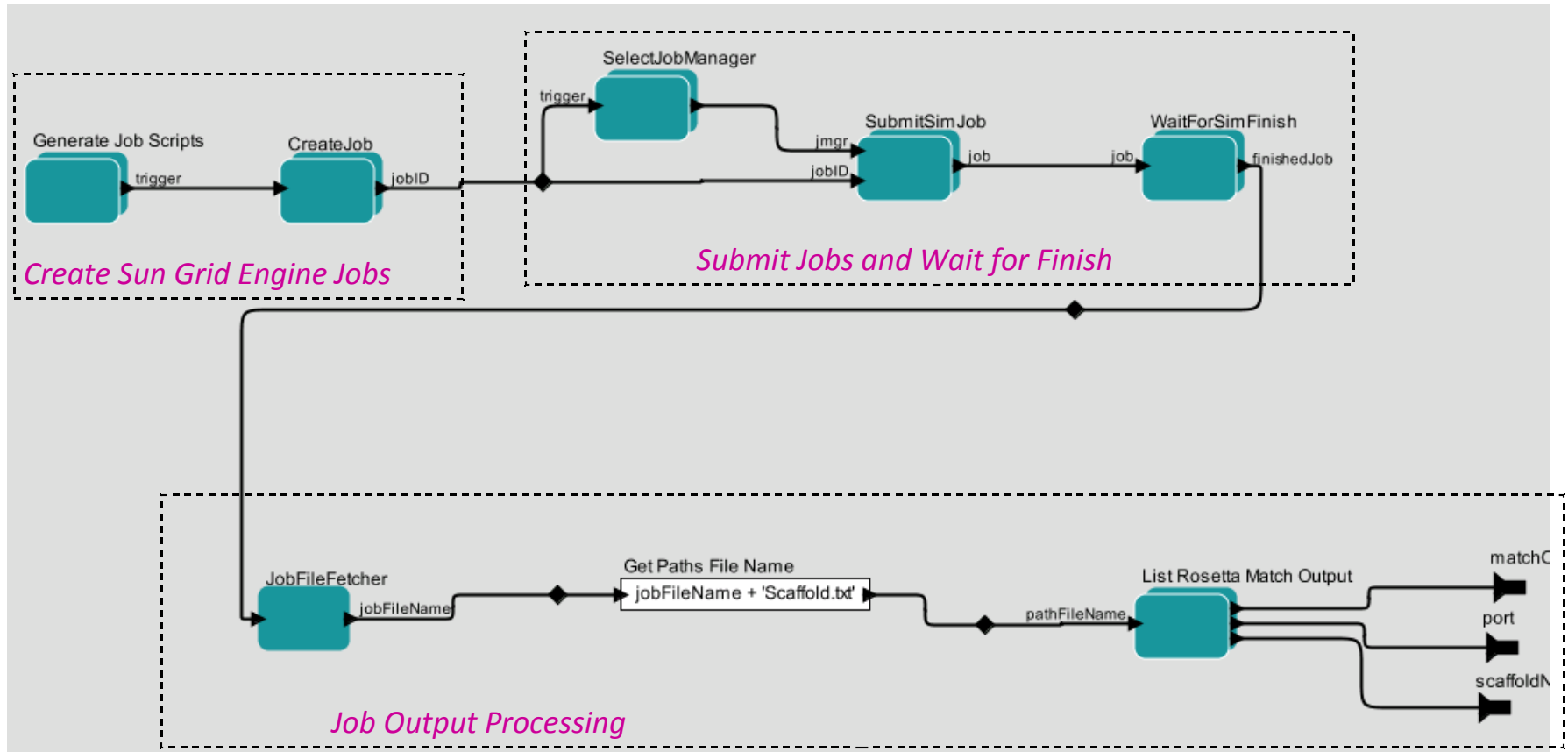


# Enzyme Design Using Kepler Workflow

- RosettaMatch
  - Goal: Finds the enzyme Scaffolds for the input theozyme
  - Approximately 250 Jobs per submission – Each creates around 2000 outputs
  - Workflow Distribute the Jobs through Scheduler/Grid
- Edge
  - Goal: Filter out Match results that are out of user defined range
  - Least time consuming – Single node job
- RosettaDesign
  - Goal: Repacks and Optimize the side chains
  - Workflow Distribute 5 million jobs from Match run through Scheduler/Grid



# Rosetta Match SGE Composite Actor



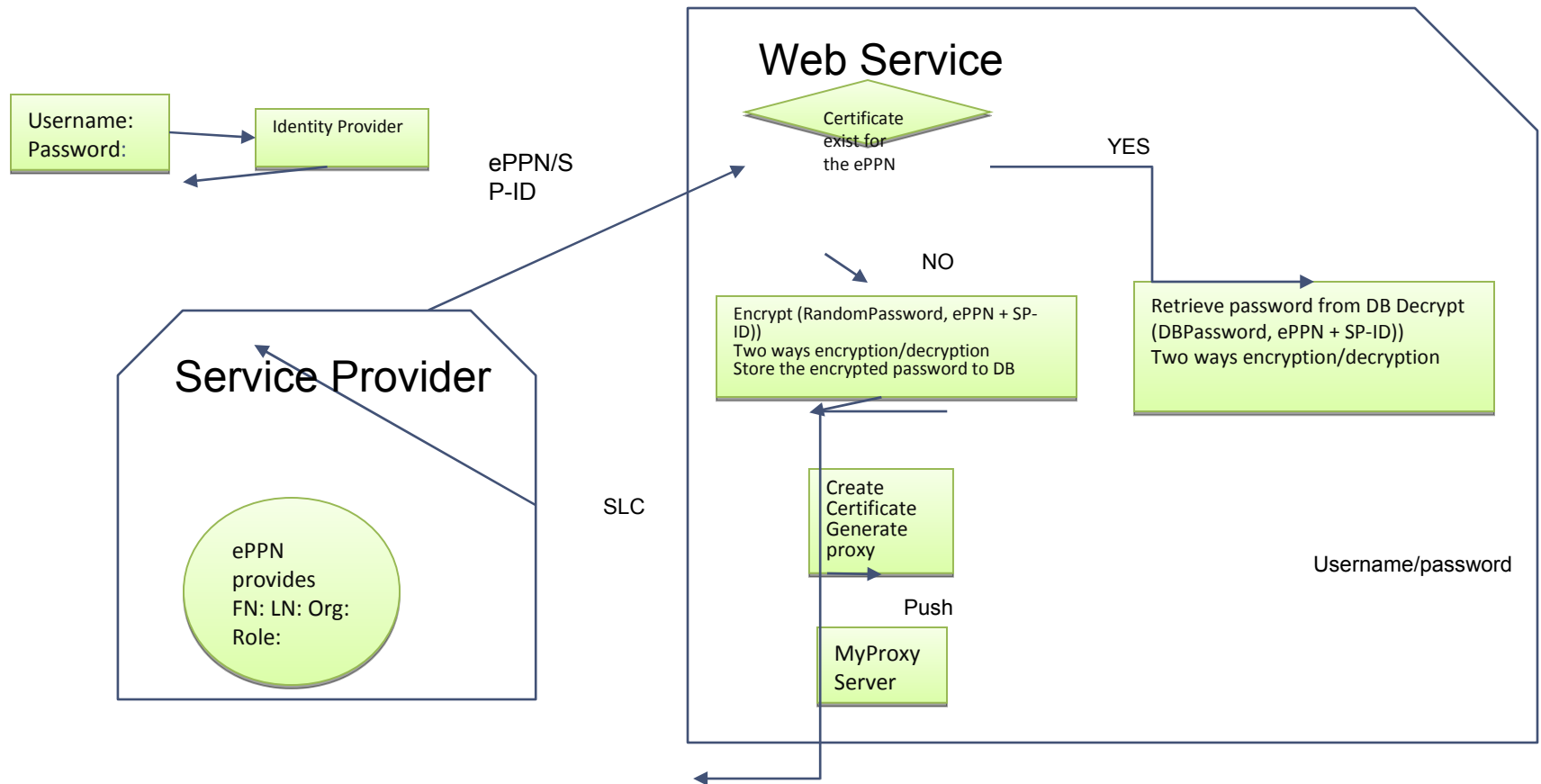
## Command line instructions for two cluster workflow (Hoffman2 and Hydro)

- Retrieve proxy from [myproxy.ucgrid.org](http://myproxy.ucgrid.org)
- `/u/local/apps/kepler/kepler.modules/build-area/kepler.sh`
  - MyProxyFilePath `/tmp/x509up_uid`
  - GlobusTaskPath4Hoffman2  
`/u/scratch/ppk/rosetta/rosetta-grid-workflow-hoffman2`
  - GlobusTaskPath4Hydro  
`/u/scratch/ppk/rosetta/rosetta-grid-workflow-hydro`
  - ScaffoldPath4Hoffman2  
`/u/home2/ppk/rosettatest/scaffold-hoffman2-one-scaffold`
  - ScaffoldPath4Hydro  
`/u/home2/ppk/rosettatest/scaffold-hydro-one-scaffold`
  - `/u/local/apps/kepler/rosetta-workflow/Rosetta-Match-BChain-Design-Globus-TopLevel-Parallel-v1-18.xml`
- TaskPaths are scratch directories for the run and ScaffoldPath is the location of input files for each cluster pre-divided by the user. Workflow directives are in the xml file

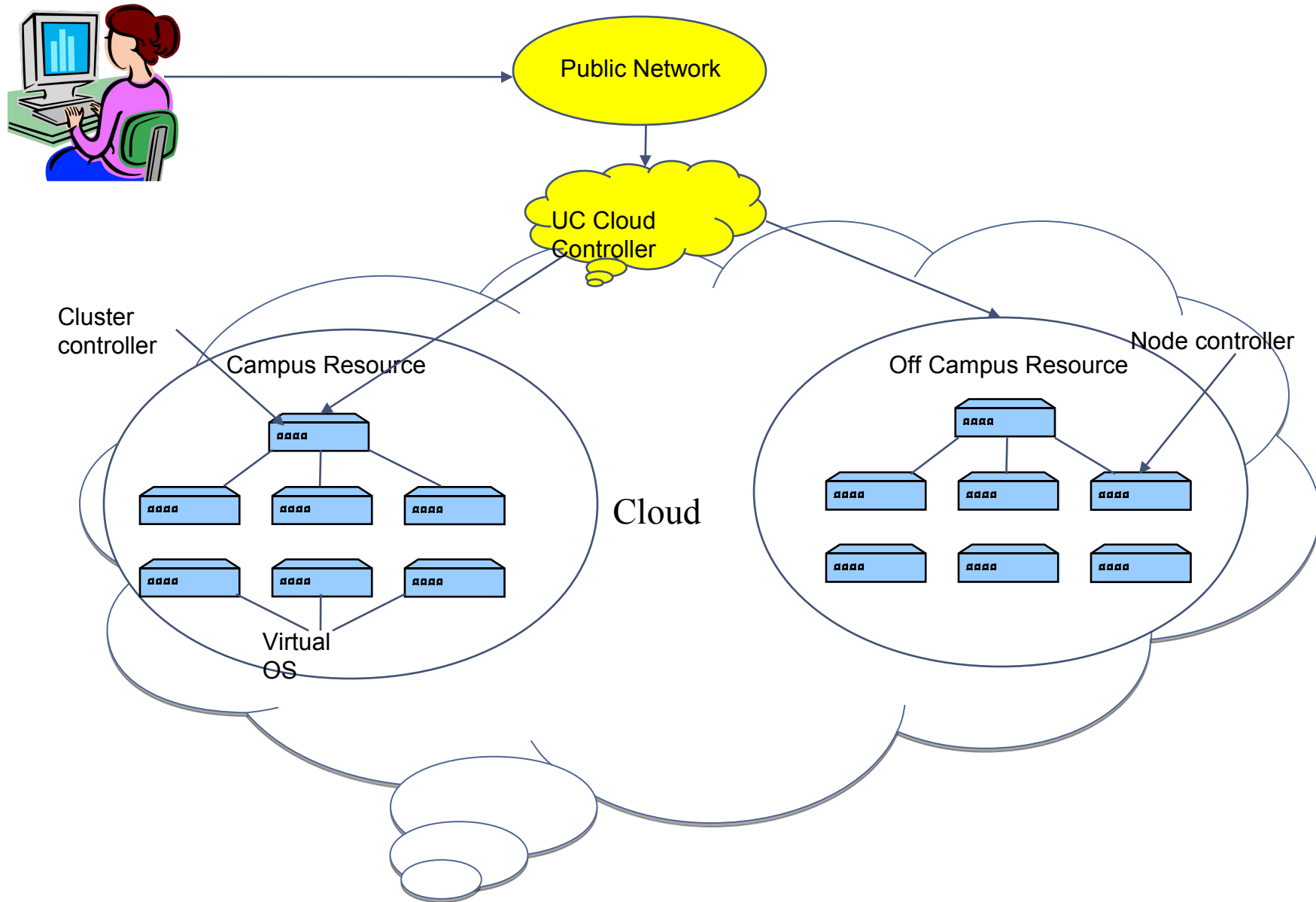
# Experimental data on Hoffman2 cluster

Workflow name	Workflow Structure	Inputs	Main configurations	Job number	Execution time
Rosetta-MatchSGE-v2-4.xml	MatchSGEJob	226 Scaffold	1) Match command number for each job: 5 Generate 108,012 pdb files.	46 Match SGE jobs.	5.36 hours
Rosetta-MatchSGE-BChain-v1-6.xml	MatchSGEJob + BChain	226 Scaffold	1) Match command number for each job: 5 Generate 216,024 pdb files.	46 Match SGE jobs.	5.19 hours
Rosetta-DesignSGE-v1-8.xml	DesignSGEJob (each design includes cst, des and min)	10 Scaffold	1) Structure Number for Design Calculation: 10. 2) Command number for each job: 5 (cst) + 5 (des) + 50 (min) = 60. Generate 14691 pdb files.	286 Design SGE jobs	10.73 hours
Rosetta-MatchSGE-BChain-DesignSGE-v1-9.xml	SGEMatch, BChain, SGEDesign	10 scaffold	1) Command number for each Match job: 1. 2) Structure Number for Design Calculation: 10. 3) Command number for each Design job: 5 (cst) + 5 (des) + 50 (min) = 60.	10 Match Jobs and 286 Design Jobs.	6.70-17.01 hours.
Rosetta-DesignSGE-v1-8.xml	DesignSGEJob (each design includes cst, des and min)	226 Scaffold	1) Structure Number for Design Calculation: 100. 2) Command number for each job: 10 (cst) + 10 (des) + 1000 (min) = 1100. The whole cmd number could be 50 million and the whole execution could generate 7 million files: * 14691 * 10 = 7,793,734,24. (2036 is the pdb file number for 10 Scaffold of Rosetta-MatchSGE-v2-4.xml)	Over 11,000 Design SGE jobs.	One job may last for over 18 hours. <b>Estimated</b> whole execution time : (108012/2036) * 10.73 * 10 = <b>7.79 months.</b>

# Workflow to Use Shibboleth Authentication to Sign X509 Certificate



# UC Cloud Architecture



# UC Grid Portal Contact Information

<http://www.ucgrid.org>

<http://inca.ucgrid.org>

[atshpc@ats.ucla.edu](mailto:atshpc@ats.ucla.edu)

- Kepler: Jianwu Wang, SDSC, [jianwu@sdsc.edu](mailto:jianwu@sdsc.edu)
- INCA, SDSC: Shava Smallen, [ssmallen@sdsc.edu](mailto:ssmallen@sdsc.edu)
- UGP: Kejian Jin [kjin@ats.ucla.edu](mailto:kjin@ats.ucla.edu)
- Rosetta: [Seonah@chem.ucla.edu](mailto:Seonah@chem.ucla.edu)
- Proteomic: [psouda@chem.ucla.edu](mailto:psouda@chem.ucla.edu)
- UCLA: Prakashan Korambath, [ppk@ats.ucla.edu](mailto:ppk@ats.ucla.edu)
- UCSB: Paul Weakliem, [weakliem@cnsi.ucsb.edu](mailto:weakliem@cnsi.ucsb.edu)
- UCR: Bill Strossman, [wlliam.strossman@ucr.edu](mailto:wlliam.strossman@ucr.edu)
- UCI: Harry Mangalam, [harry.mangalam@uci.edu](mailto:harry.mangalam@uci.edu)