THE UNIVERSITY OF CHICAGO | Center for Research Informatics

WORKSHOP SERIES

Using CRI's Computational Infrastructure for Bioinformatics Analysis

Instructors: Mike Jarsulic, Olumide Kehinde, Kyle Hernandez

CRI SYSTEMS RESOURCES OVERVIEW

HPC

Tarbell Cluster: High Performance Computation Cluster for Data Analysis

Large Memory Server: (SAS/STATA/X-Windows)

Web Portals

Galaxy: Integrates with the HPC & provides a suite of tools for data analysis

REDCap: for building online surveys and databases. Research data can be exported to statistical packages (SPSS, SAS, Stata, R)

Data Storage

Home Data: 10GB Limit

Lab Shares: From 100GB

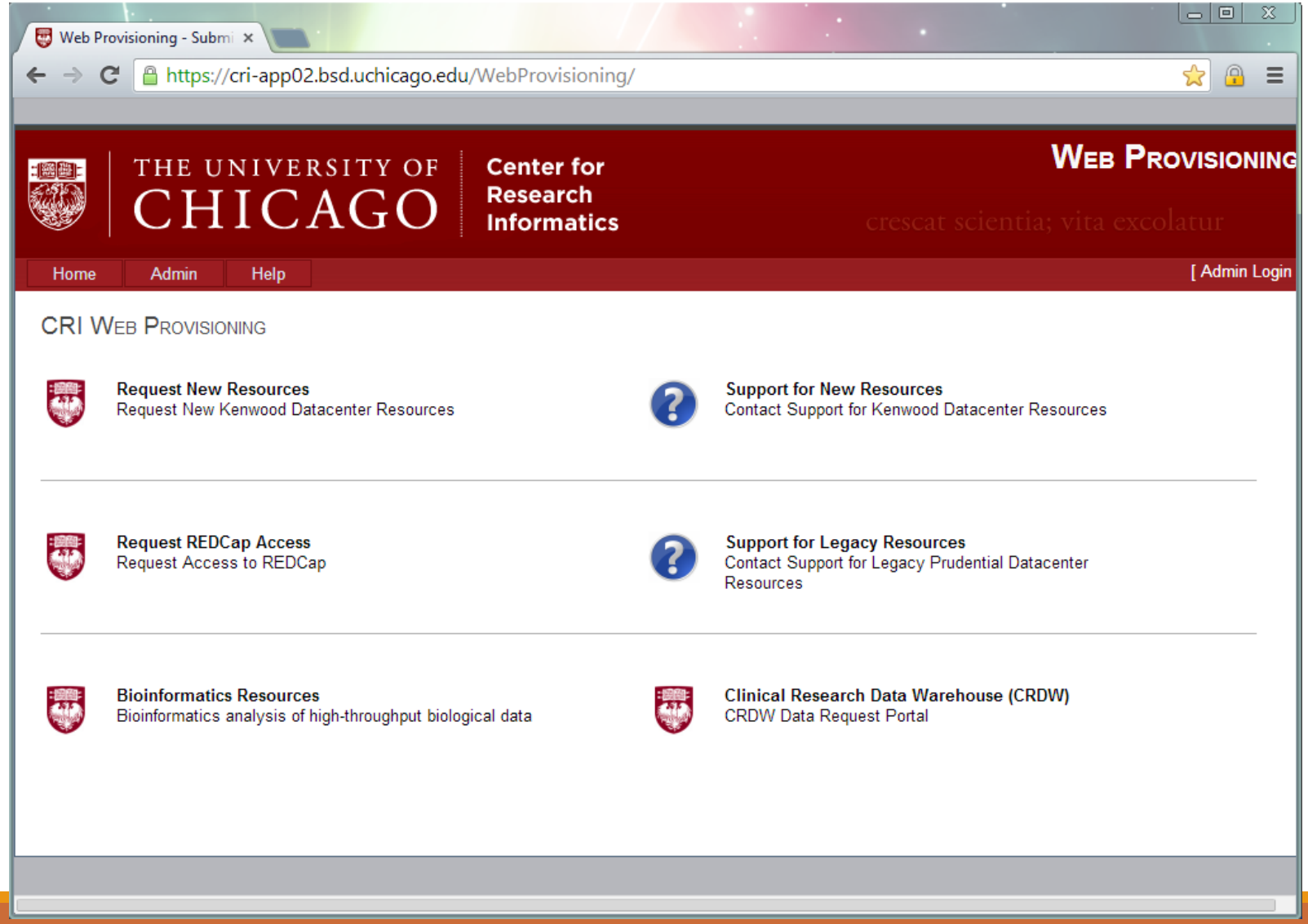
Fast Scratch Space: Shared 60TB

Virtualization

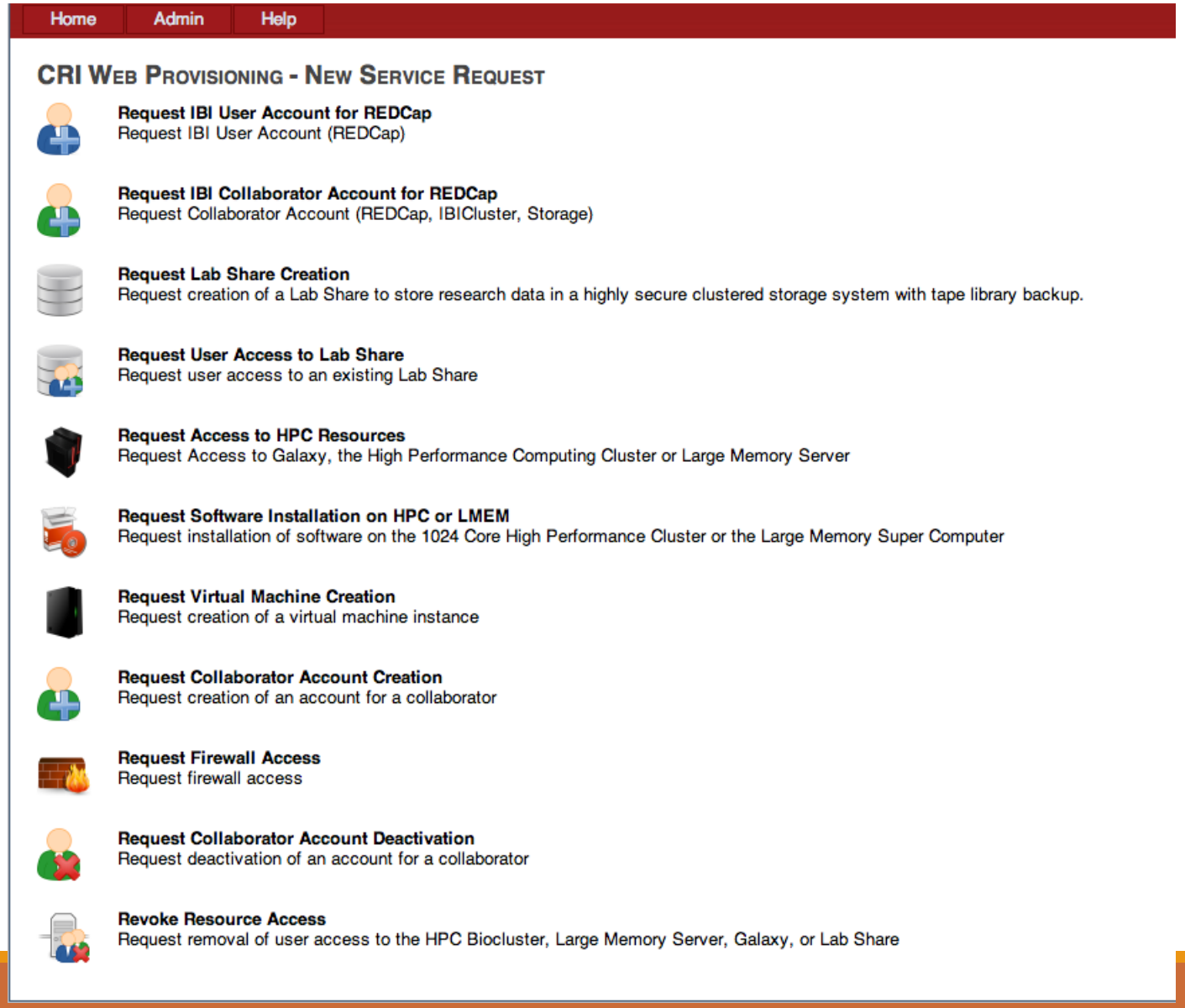
Virtual Machines: Hosted Windows & Linux Servers for building web applications and databases

HOW TO REQUEST CRI RESOURCES/SUPPORT

Visit: **cri.uchicago.edu > Request Resources**



ONLINE FORMS



REQUESTING SUPPORT

1) For BSD Account issues, please contact CBIS help@bsd.uchicago.edu or call 773-702-3456

2) Use appropriate form on the Request Site to create a support ticket;
   Visit: cri.uchicago.edu, Request Resources Link
   -Trouble with installed software/ modules
   -Job submission issues
   -Questions

1) Call the CRI Help Desk to follow up on tickets - 773-834-8475

DATA MANAGEMENT

-Home Directories: /home/<username>
Limited to 10GB. Single user access to data
Check your current working directory with command #pwd

-Lab Shares: /group/<share-name>
Group access to backed-up data
Changing directory: #cd /group/share-name

-Scratch Space: /scratch/<share-name>
60TB Space for staging input used in analysis jobs and for temp files from job execution.
Data in scratch is purged periodically.

-Remote Access to Data
Login to the Campus VPN (cvpn.uchicago.edu)

DATA TRANSFER TO/FROM CRI STORAGE

Accessing Share From a Windows PC:

Option 1:

Click the Start Button > Type "run" in the Search Box and press "Enter" > In the run window, type \\bulkstorage.uchicago.edu\<Lab-share-name>, press "Enter"

Login with your BSD Account, format "BSDAD\<bsd-username>"

Option 2:

From "My Computer" > "Map Network Drive"

Folder name: \\bulkstorage.uchicago.edu\<Lab-share-name>

*If you are not logged in to the BSDAD Domain, Check "Connect using different credentials"

Login with your BSDAD Account, format "BSDAD\<bsd-username>“

Accessing Share From a Mac:

From "Finder" > "Go"

Server Address: smb://bulkstorage.uchicago.edu/<Lab-share-name>

Login with your BSDAD Account, format "BSDAD\<bsd-username>"

DATA TRANSFER TO/FROM CRI STORAGE (Cont'd)

Using Rsync, SCP or SFTP to transfer large data:

Host: cri-syncmon.cri.uchicago.edu
Login: BSD Account/Password
Path: /group/<share-name>
Port: 22

Examples of Free SFTP Graphical Clients:

WinSCP, CoreFTP (Windows Only) or
FileZilla, Cyberduck (Windows & Mac)

HOW TO USE ANALYSIS RESOURCES

High Performance Computing (HPC) Cluster:

Execute analysis in batch mode

Execute analysis in interactive mode

X Window System

Large Memory Server:

Execute SAS or Stata analysis

X Window System

SPECIFICATIONS

Interactive Nodes:

2 Nodes
2 AMD Opteron 6274 processors (32 cores)
128 GB RAM
Scientific Linux 6 Update 4

Compute Nodes:

34 Nodes
4 AMD Opteron 6274 processors (64 cores)
256 GB RAM
Scientific Linux 6 Update 4

SPECIFICATIONS (Cont'd)

Large Compute Nodes:

2 Nodes
4 Intel Xeon E7-4860 processors (40 cores)
Hyper-threaded
1 TB RAM
Scientific Linux 6 Update 4

Large Memory Server:

8 Intel Xeon E7-8870 processors (80 cores)
Hyper-threaded
1 TB RAM
Redhat Linux 6 Update 4

PRODUCTION ENVIRONMENT

Complete:

- TORQUE Resource Manager
- MOAB Scheduler
- Galaxy
- Scratch Space (60 TB)

Coming Soon:

- Web services
- Simplified Parametric Submission
- Job portability
- Trickled submission
- More Scratch Space

SOFTWARE ENVIRONMENT

Compilers:

- GNU Compilers
- Intel Compilers
- PGI Compilers

Language Support:

- C/C++
- FORTRAN
- Perl
- Python
- R

BEST PRACTICES

- DO NOT – Install Software into your home directory/lab share
- DO NOT – Write your job scripts using MS Word
- DO NOT – Copy/Paste from MS Word documents into your jobs
- DO – Contact ME (mjarsulic@bsd.uchicago.edu) if you are planning on running a large analysis
- DO – Understand that Java programs grow in memory
- DO NOT – Run analysis on the interactive nodes