Remote Research Computing

Contact: researchcomputing@rice.edu
Visit our new site!! http://ResearchComputing.rice.edu
Updated Knowledge Base Documentation, and new CRC website!

The Center for Research Computing (CRC) works to meet Rice community members' research needs with specialized computing resources and services. We maintain shared research computing infrastructure on campus as well as relationships with off-campus organizations and vendors, and facilitate the optimal use of these resources through consultation and direct partnership.
CRC Resources

And Connection Methods
Services and Resources

• RDF Data Storage, NOTS Cluster, ORION Private Cloud, GLOBUS FTP
• Clusters, storage, and VMs can all be accessed via VPN.
• Documentation available on Knowledge Base: https://kb.rice.edu/108191
• Our facilitators, application support specialists, developers, and staff are available to help you:
  • Access resources at Rice and elsewhere
  • Design a remote-access workflow
  • Run workshops tailored to your group’s needs
CRC Networked Resources

Solid lines represent internal Rice or I2 connections
Dotted lines represent commercial internet connections
Remote Access Methods for Optimizing Workflows

Firewall:
- SSH gateways to cluster
- HTTPS access to Rice VM’s with public IP addresses/Netscaler

VPN
- With Duo authentication
- Direct
- Research networks to Rice Data Transfer Nodes (Science DMZ)

https://kb.rice.edu/108626
Rice’s Research Data Facility
The Research Data Facility

- Networked Isilon storage
- Automated backups
- Shares are available for research faculty
- Mount on a VM or your laptop as SMB share
- Direct connection to research networks via Science DMZ

https://kb.rice.edu/108241
RDF EXERCISE

• CONNECT TO VPN
• Connect to RDF
• Copy a file to RDF using Terminal or GUI
VPN Examples
Connect via VPN

Solid lines represent internal Rice or I2 connections
Dotted lines represent commercial internet connections

See: https://kb.rice.edu/82263
Supercomputing and Storage
Supercomputers and Storage: Staging Data

For large datasets, do:
1. Use Globus to ship the data to NOTS
2. Execute the job on NOTS
3. Ship the data back to the Isilon

For medium datasets, you can:
1. SMB mount a VM on the RDF
2. Use the VM to SCP data to & from NOTS

For small datasets, you can:
1. SCP the data to & from nots via vpn
Supercomputers and Storage: Staging Data

For large datasets, do:
1. Use Globus to ship the data to NOTS
2. Execute the job on NOTS
3. Ship the data back to the Isilon

For medium datasets, you can:
1. SMB mount a VM on the RDF
2. Use the VM to SCP data to & from NOTS

For small datasets, you can:
1. SCP the data to & from nots via vpn
Supercomputers and Storage: Staging Data

For large datasets, do:
1. Use Globus to ship the data to NOTS
2. Execute the job on NOTS
3. Ship the data back to the Isilon

For medium datasets, you can:
1. SMB mount a VM on the RDF
2. Use the VM to SCP data to & from NOTS

For small datasets, you can:
1. SCP the data to & from nots via vpn
Globus’ File Manager Web Interface

Using Globus’ web interface to:

• Browse my personal NOTS scratch directory (/scratch/jcm10/)
• Create a destination folder on my RDF/Isilon share (/research/jcm10/)
• Initiate the transfer of an entire subdirectory, containing ~150,000 files in ~800 directories.

https://app.globus.org
Globus vs. VM-scp vs. VPN-scp Dataset Example

Test transfer, NOTS to RDF
- Statistics
  - 1875 files
  - 651 directories
  - 50GB
- Globus (unencrypted)
  - Without file verification
    - 156 MB/s
    - 5 minutes
  - With file verification
    - 79 MB/s
    - 11 minutes
- SCP to campus Linux desktop, ethernet (not wifi)
  - 107 minutes
  - No file verification

Log in at [https://www.globus.org/](https://www.globus.org/) with your Rice ID
Instructions for use on KB: [https://kb.rice.edu/108242](https://kb.rice.edu/108242)
DATA xFER EXERCISE

• GLOBUS
  • Log into Globus FTP
  • Connect your endpoints:
    • NOTS scratch directory
    • RDF folder

• SCP
  • Connect to VPN
  • Connect to your RDF share
  • Open your terminal
  • SCP the file from NOTS to your RDF share
Globus’ Command Line Interface

Using Globus’ CLI to automate tasks
• Create event listeners & scripted tasks (such as, when a NOTS job finishes, export files to RDF)
• Duplicate transfers (such as exporting data from NOTS to RDF and external repositories)

https://github.com/globus
ORION Virtual Machines
ORION Virtual Machines

Web interface to manage VM’s

- Small (2 cores x 8 GB RAM), medium (4x16), large (8x32)
- SMB fileshare mounting
- Fast connection on Rice network
- Can provision public IP’s or Netscaler proxy hostname
- A variety of disk images are available

Orion Documentation

https://orion.crc.rice.edu

https://kb.rice.edu/108243
Workhorse VM with Mounted Storage

- Pre-processing data for clusters
- Post-processing data for clusters
- Medium-sized jobs
- Interactive jobs
Lightweight VM with Mounted Storage, Private IP

Data processing example:
Database migration via API calls:
- Didn’t complete overnight on my laptop via VPN
- 4.1 seconds per item
- Moved code onto target ORION VM
- 0.43 seconds per item

Data Interface Example:
Lightweight Flask app connecting to RDF datasets (courtesy of Kavrakilab: Anja Conev, Nonso Chukwurah, Romanos Fasoulis)
NETWORKING EXERCISE

• Create an SSH key
• Launch an ORION VM
• SSH in
• Connect to networked storage
• SCP data from NOTS to RDF *without* touching down in your own computer
Wrapping up
Rice research computing resources are

• Fast
  • Connections up to 100MB/s
  • Scalable storage and computing

• Safe
  • Firewalled
  • Authenticated users only
  • Reliably available
  • Less subject to demands on the commercial internet
  • Predictable pricing

Contacts:
• researchcomputing@rice.edu
• helpdesk@rice.edu
• Facilitators:
  • Clinton Heider: heider@rice.edu
  • John Mulligan: jcm10@rice.edu