# Remote Research Computing

Contact: researchcomputing@rice.edu

Visit our new site!! <u>http://ResearchComputing.rice.edu</u>

# Updated Knowledge Base Documentation, and new CRC website!

	CRC Remote Access to Service	× +				
←) → ⊂	🗊 🔒 https://kb.	rice.edu/108626			ז 🚽 🛙	\ □ ◎ ∅ # =
💖 sed gnu reference	e () .md 💁 OUTLOOK (12)	😝 AWS 📋 crc site login 📋	] New Folder 🛛 🛅 python	PYPI 🁲 Trav	ris tutuorial 🛐 crc	login » 🛅 Other Bookmark
Remote Teaching	Teaching Continuity Remote	Learning Remote Working			KB Home	OIT Rice University
	<b>N</b> DIOT					
	RICE	Search the KB		All Topics	0	SEARCH Advanced
ī	Teaching	Research	Administr	ation	IT Sec	curity
			Get Help			
IT Status V KB Home	CRC Remote Ac	cess to Services a	nd Resources			
Adobe Service Status	CRC Remote Access to Se	ervices and Resources				Highlight search terms
Box Status	Off-Campus Acc	ess to CRC Servic	es and Resour	ces		
Canvas Service Status	All the primary computing s	systems & services offered by intermethods of remote access	CRC are available to res	searchers who are	working from off ca	mpus. Please see the
Google Status						
Kaltura System Status	CRC Service		Off-Campus Access	Method		
Zoom Service Status	High Performance Cluste	er(s) (NOTS)	Rice VPN or CRC ss	h gateway required		
Quick Login Links 🗸	Research Data Facility -	Isilon (RDF)	Rice VPN required			
Webmail   Email Web Access	ORION Virtual Machines	for Research - Web Console	Rice VPN required			
Rice Google Mail	REDCap		Public access (DUO	required)		
Microsoft 365 Exchange Online	CRC Documentation Public access					
Outlook Web Access (Exchange Mail Users	GLOBUS file transfer Public access (NetID authentication)					
Only)	VRDE Secure Research	Enclave	CEHI2FA VPN requi	red		
Google Calendar Login	Consultation/Facilitation		Can be arranged via	Zoom or Google		
Office 365	<u> </u>		Weet			
Canvas Learning Management System	In Annil 2020 Alta OCC. "	and have been as the second				
Zoom	in April 2020, the CRC Will	start nosting webinars with a fe	Jous on remote researc	in computing practic		



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.

#### https://kb.rice.edu/108626

#### https://researchcomputing.rice.edu/

# 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: <u>https://kb.rice.edu/108191</u>
- 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



### 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



•••	🖏 CRC Research Data Fac	illity × +						
← → ℃ ·	🗟 🔽 🔒 https://	kb.rice.edu/108241		🗵 🖒	• 👱 🖩	1 🗉 🔮	6)	₫ Ξ
💖 sed gnu refe	rence 🔘 .md 💁 OUTLO	DOK (12) 🧉 AWS 🛅	] crc site login	🗎 New Folder	python	» [	Other	Bookmarks
Remote Teaching Learning Remot	J Teaching Continuity e Working	Remote			KB Home	e OIT I	Rice Unive	ersity
	<b>RICE</b>	Search the KB	9		All Topics			SEAI
	Teaching	Res	earch		Administrat	tion		
		IT Security		Get Help				
IT Status V KB Home	CRC Research	Data Facility	,					
IT Service Status	CRC Research Data Fa	cility				Highligh	nt search t	terms
Adobe Service Status	To accommodate our research community's growing need to store large datasets and facilitate collaboration among research groups, the Center for Research Computing has made available a variety of services collectively known as the Research Data Facility (RDF). The RDF consists of a combination of cloud-based and on-premises storage services that are robust and secure, flexible enough to meet a wide range of use cases, and scalable to meet future data storage needs. Security							
Box Status								
Canvas Service Status								
Google Status Kaltura System Status								
Zoom Service Status	The RDF is designed to securely accommodate non-regulated data. If your data requires additional security precautions to ensure regulatory compliance, please contact the CRC to discuss your requirements.							
Quick Login Links V Webmail   Email Web Access	Rice Box			- Heberstine		detects -		
Rice Google Mail	Many researchers require a secure data archive or collaborative space where datasets can be shared easily with colleagues both inside and outside of Rice, but there is not a need for direct, real-time access to the data from software applications. The Rice Box service is designed to meet those							
Microsoft 365 Exchange Online	requirements. R A cloud s	tice Box is:	ed inside the l	JS				
Outlook Web Access (Exchange Mail Users Only)	<ul> <li>A cloud storage sourch of asked mixed the US</li> <li>Accessible via your Rice NetID and password</li> <li>Allows files/folders to be easily shared to Rice colleagues</li> <li>Allows links to files to be sent to external collaborators' email addresses</li> <li>Supports FTP bulk data transfer</li> <li>Currently, limited to commercial internet speeds – the CRC recommends consulting with us about integrating cloud storage solutions into workflows or archiving practices that involve</li> </ul>							
Google Calendar Login								
Office 365	Can be s	ynchronized to loca	I folders				00000	
Canvas Learning Management System	Limited s To get started, vis	torage (Please Note sit Getting Started w	the change f	rom Unlimited	here https://k	(b.rice.edu/	69392)	

#### https://kb.rice.edu/108241

- 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

#### RDF EXERCISE

- CONNECT TO VPN
- Connect to RDF
- Copy a file to RDF using Terminal or GUI

# VPN Examples





#### See: https://kb.rice.edu/82263

Dotted lines represent commercial internet connections

# 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

Image: Second state	manager?destination_id=5416 Projects - Google S 💽 MDN ≶	3 SQLite 👩 SIEGE 🗎 RT 🛂 IU I	MAIL ScRC chat Oceans and Contine.		> 🤎 🔟 🤘
🗂 File Manager				Pa	anels
Collection NOTS scratch		Q 🛞	Rice University CRC RDF end	lpoint	Q (8)
Path /jcm10/			/rdf/jcm10/riedeltest/		
select none 🗠 🖒		kĝis view →Ξ	select none 1 $C$		ېن {ڼ
NAME ~	LAST MODIFIED	Share 6	NAME 🗸	LAST MODIFIED	SIZE
files	03/25/2020 11:0	Transfer or Sync to	riedel_jan7.tar.gz	01/07/2020 10:	19.06 MB
gdal2tiles-leaflet-master	02/19/2020 02:4	New Folder	run notes.txt	12/02/2019 02:	1.29 KB
master.zip	03/26/2020 10:0	Rename 💋	SanbornBlocks	11/05/2019 04:	*
slurm-532942.out	<i>03/27/2020 06:</i> C	Delete Selected 🔀	SanbornLots	11/06/2019 04:	-
tilemaker2.slurm	03/26/2020 10:0	Download 다. Open 12	Tiles	11/20/2019 10:	-
tilemaker3.slurm	03/25/2020 05:2	Upload 4	tilesout-march25	03/26/2020 12	
tiles	03/26/2020 10:0	Get Link @	tilesout-march26	03/26/2020 09	-
		Show Hidden Items	tilesout-march26a	03/26/2020 06	
		Manage Activation	tilesout-march27	03/27/2020 02	
				0.0	

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 <u>https://www.globus.org/</u> with your Rice ID Instructions for use on KB: <u>https://kb.rice.edu/108242</u>

### 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

O glob	bus/globus-cli: A Command 🗙 🕂					
← → ♂ ŵ	🛛 🔒 https://github.com/globus/globus-cl		··· 🛛 1	Search		lii\ 🦑 💿 🗊 🛎 😑
AWS OED 🖏 JC	M research 🌐 https://mpolson64 🚹 Projects - G	oogle S 💽 MDN 🚫 SQLite 🚺 SIEGE 🗎	🗒 RT 📴 IU MAIL 🥥 CRC chat 🛛	Oceans and Contine		
Search or jum	p to Z Pull requests	Issues Marketplace Explore				🖍 ++ 📵+
	🖟 globus / <b>globus-cli</b>		Øv	Vatch + 14 🕈 Star	31 ¥ Fork 12	
	<> Code ① Issues 16 ① Pt	Il requests 3 🔘 Actions 🔟 Proje	ects o 📧 Wiki 🕕 Secur	rity de Insights		
	A Command Line Wrapper over	the Globus SDK for Python				
	-> 827 commits § 7	vranches 🗇 0 packages	S 38 releases	6 contributors	₫ Apache-2.0	
	Branch: master - New pull reques		Create new file	Upload files Find file	Clone or download -	
	jaswilli Merge pull request #487 f	rom globus/release-1.12.0		🗸 Latest commit	764d9d7 15 days ago	
	adoc	Update developer tools			2 months ago	
	🖬 globus_cli	Release v1.12.0			15 days ago	
	shell_completion	Replace completer script	Replace completer scripts with 'completer' opts			
	in tests	Remove test case for failu	ure on click<7.1		16 days ago	
	.coveragerc	Update developer tools			2 months ago	
	.editorconfig	Add .editorconfig			4 years ago	
	.gitignore	Update developer tools			2 months ago	
	.pre-commit-config.yaml	Update developer tools			2 months ago	
	.travis.yml	Update Travis to not run l	linting on each platform		2 months ago	
	CONTRIBUTING.adoc	Update developer tools			2 months ago	
https://mpolson64.github.io/O	DceansAndContinentsTESTBED/dist/	Initial commit			4 years ago	

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

		Not officially supported
reate Virtual Mach	ine	
r Matlab VM	Pers	istent 🖗
		Create
mplate		
matlab ALL	🔹 🚯 Labels 💌	
MATLAB 2017b - GUI - Small	MATLAB 2017b - GUI - Medium	MATLAB 2017b - GUI - Large
	🔅	45
	CentOS	CentOS CentOS 7 clouid image version
CentOS 7 cloud image version	CentOS 7 cloud image version	South Start Start Start Start Start
CentOS CentOS 7 cloud image version 1802 with development tools,	CentOS 7 cloud image version 1802 with development tools,	1802 with development tools,
CentOS CentOS 7 cloud image version 1802 with development tools, GNOME, and MATLAB 2017b	CentOS 7 cloud image version 1802 with development tools, GNOME, and MATLAB 2017b	1802 with development tools, GNOME, and MATLAB 2017b
Centors Centors 7 cloud image version 1802 with development tools, GNOME, and MATLAB 2017b Small machine size	CentOS 7 cloud image version 1802 with development tools, GNOME, and MATLAB 2017b Medium machine size	1802 with development tools, GNOME, and MATLAB 2017b Large machine size

#### https://orion.crc.rice.edu

#### **Orion Documentation**

	CRC Getting started on OR	ION X +					
€ → C @	🛛 🏔 https:	//kb rice.edu/108243	E	🖂 🟠	1 IN II I II I		
👘 sed gnu reference	🔿 md 💁 OUTLOOK	( (12) 🧯 AWS 🛅 crc site login 🛅	New Folder 🛅 python	PYPI 🔮 Travis tutuori	ial 🚺 crc login ≫ 🛅 Other Bookmark		
Remote Teaching	Teaching Continuity Re	mote Learning Remote Working			KB Home OIT Rice University		
	DICE						
	GIUCL	Search the KB		All Topics	SEARCH Advant		
	Teaching	Research	Administratio	94	IT Security		
		0	et Help		1.2		
TT Status 👻	0000						
KB Home	CRC Getting	started on OkiON					
Adobe Service Status	Getting started on OR	ION CRC guide			Highlight search terms		
Bex Status	ORION						
Canvas Service	Owl Research Infras	itructure Open Nebula (ORION) V	M Pool on Rice Campus	is, in essence, a linux	virtual machine farm.		
Status	ORION offerings are	comprised of:					
Kaltura System Status	<ul> <li>Virtual Machin interface once</li> </ul>	es of various Linux OS' and softw you've been given access.	are packages. Please cl	heck our catalog for ou	ir offerings via the ORION web		
Zoom Service Status	We offer the followin	We offer the following Virtual Machine support levels:					
Quick Login Links 🗸	1. Full admin: A 2. Customer ad	n IT Systems Administrator takes min: IT assumes no responsibility	full responsibility for sys for stability of the opera	tems administration of ting system.	f the VM.		
Webmail   Email Web	To obtain access to	ORION please apply for access as	you would for any othe	r CRC computational r	resource and select ORION as		
Rice Google Mail	or via VPN and conr	your account has been approved sect to: https://orion.rice.edu/ via y	by your sponsor you can our netID and netID pase	sword. The web interfa	web browser either on campus ace is not accessible directly from		
Microsoft 365	the internet.						
Exchange Online	Please note we are a	available for discussions if our cur	rent offerings don't meet	your needs. We may	be able to accommodate you.		
(Exchange Mail Users Only)	Prease contact us in	a bower and relefence Oktow.					
Google Calendar Login	LOGIN to V	M on ORION					
Office 365	Typically when you h	ave instantiated a VM it will be ac	cessible via your SSH k	ey and your netID.			
Canvas Learning Management System	To ensure your SSH operating system yo	key is present on your VM you m u are using e.g. Linux, Mac, or Wi	ust have generated a SS ndows.	iH public and private k	ey. This will depend on the		
Zoom	1. Login to ORIC	N using your web browser using	your netID and netID par	ssword.			

#### https://kb.rice.edu/108243

- 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

### Workhorse VM with Mounted Storage



### Lightweight VM with Mounted Storage, Private IP



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

### Research Computing Remote Access



#### 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
- <u>helpdesk@rice.edu</u>
- Facilitators:
  - Clinton Heider: <u>heider@rice.edu</u>
  - John Mulligan: jcm10@rice.edu