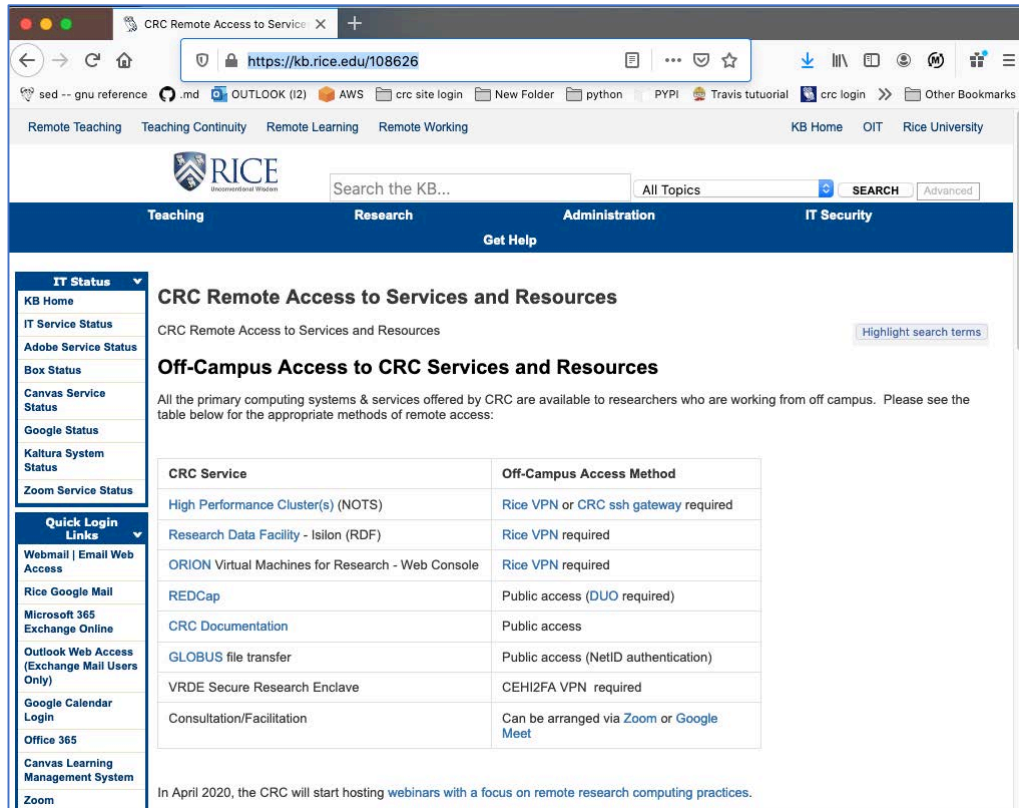


Remote Research Computing

Contact: researchcomputing@rice.edu

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

Updated Knowledge Base Documentation, and new CRC website!



The screenshot shows a web browser window with the URL <https://kb.rice.edu/108626>. The page features a navigation bar with links for Teaching, Research, Administration, and IT Security. A search bar is present with the text "Search the KB...". The main content area displays the title "CRC Remote Access to Services and Resources" and a sub-section "Off-Campus Access to CRC Services and Resources". A table lists various CRC services and their corresponding off-campus access methods.

| CRC Service | Off-Campus Access Method |
|---|---|
| High Performance Cluster(s) (NOTS) | Rice VPN or CRC ssh gateway required |
| Research Data Facility - Isilon (RDF) | Rice VPN required |
| ORION Virtual Machines for Research - Web Console | Rice VPN required |
| REDCap | Public access (DUO required) |
| CRC Documentation | Public access |
| GLOBUS file transfer | Public access (NetID authentication) |
| VRDE Secure Research Enclave | CEH2FA VPN required |
| Consultation/Facilitation | Can be arranged via Zoom or Google Meet |



The screenshot shows the homepage of the Center for Research Computing (CRC) at Rice University. The header includes the Rice University logo and the text "RICE UNIVERSITY Center For Research Computing". A navigation menu is visible in the top right corner. The main content area features a large image of server racks with the text "The Center for Research Computing" overlaid. Below the image, a paragraph describes the CRC's mission: "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."

<https://kb.rice.edu/108626>

<https://researchcomputing.rice.edu/>

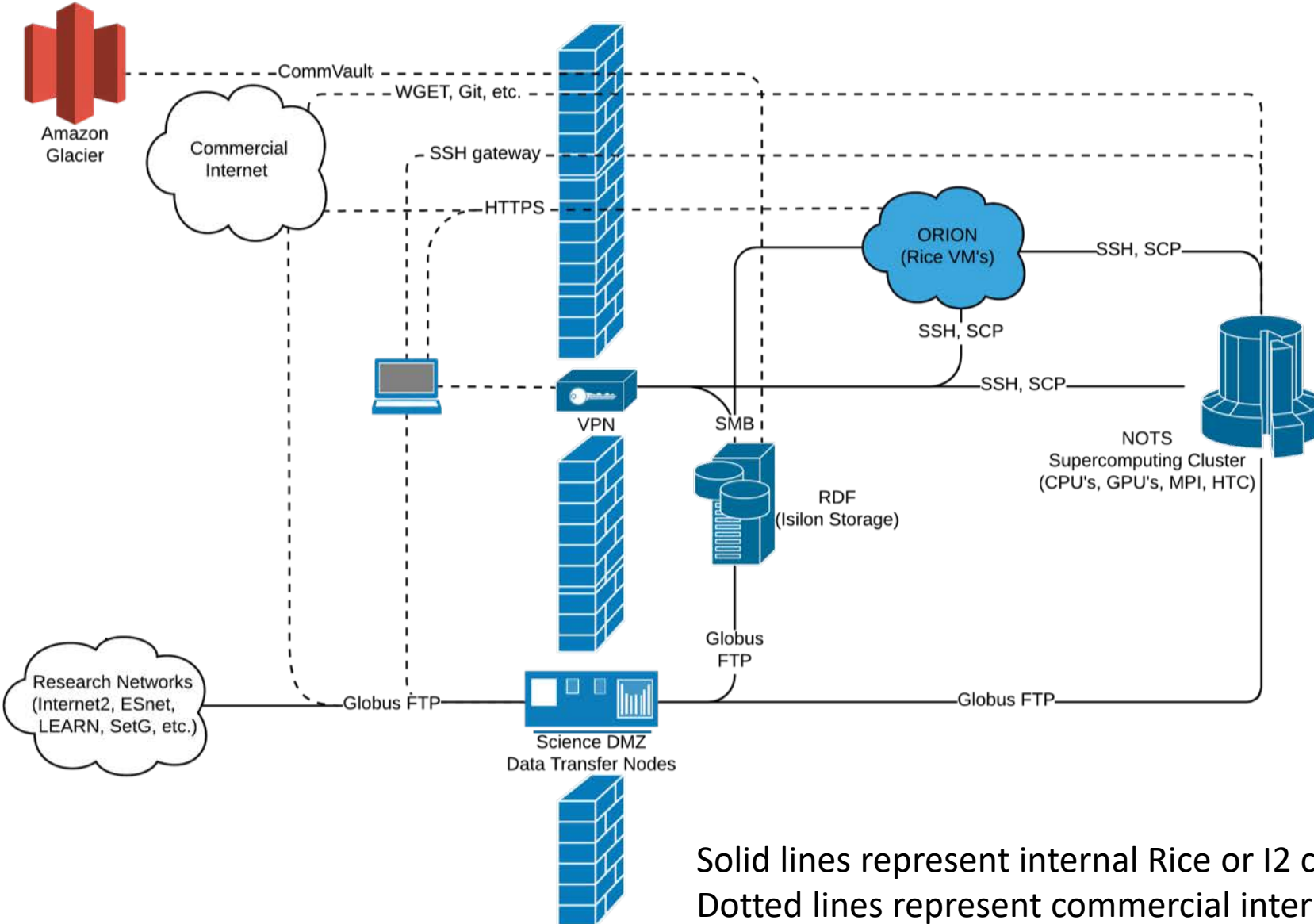
CRC Resources

And Connection Methods

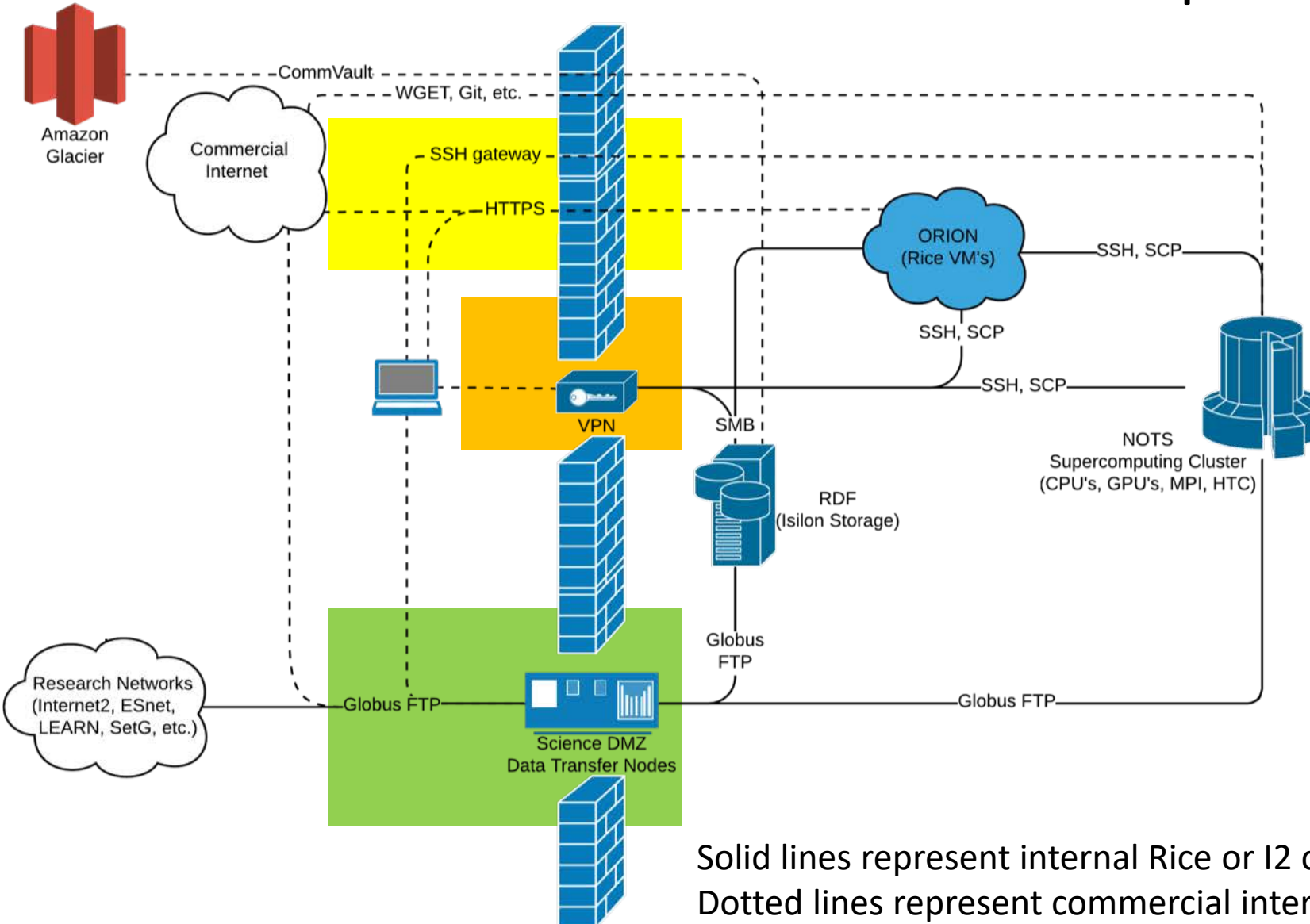
Services and Resources

- 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:
<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



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

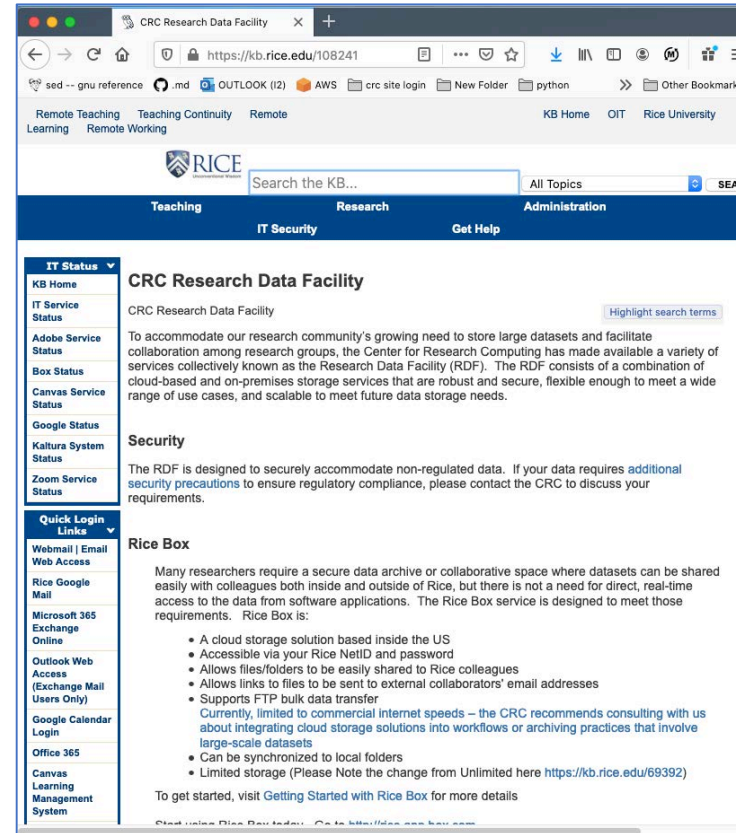
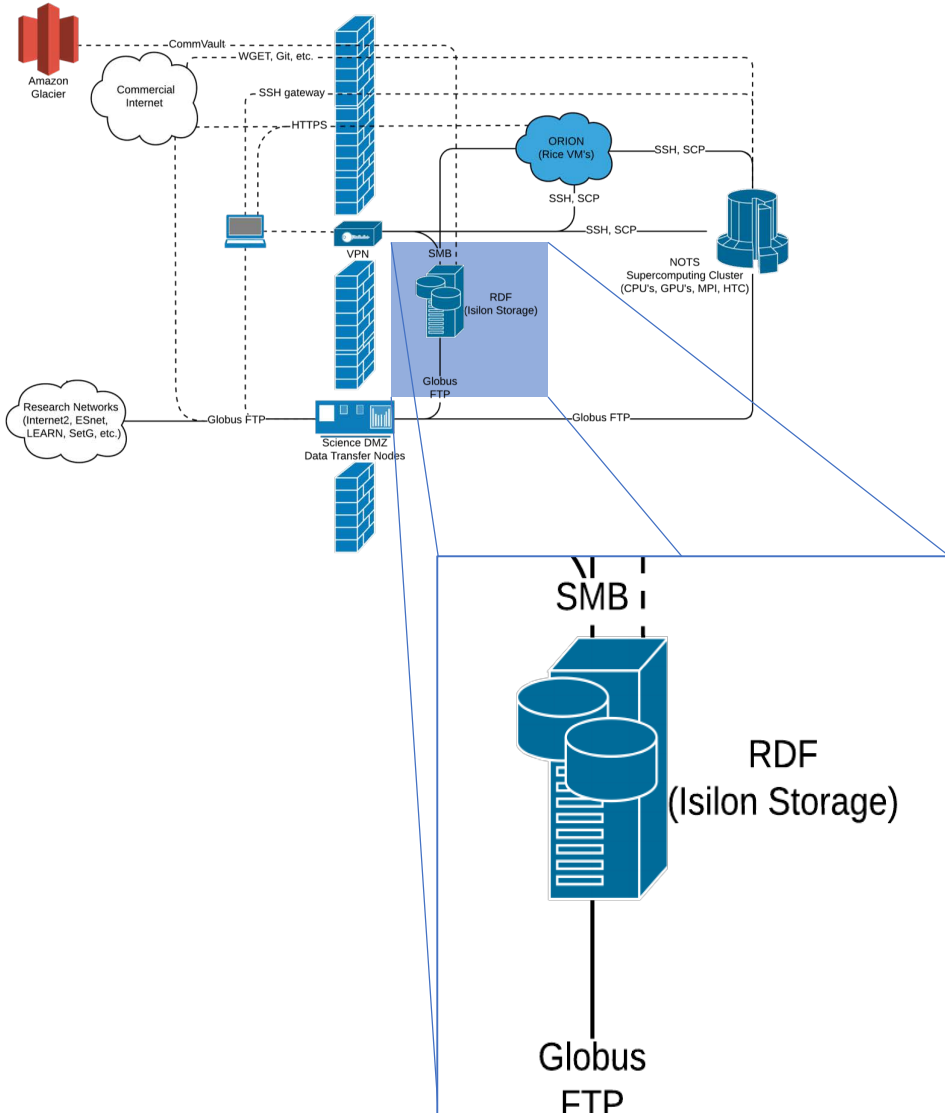
- Research networks to Rice Data Transfer Nodes (Science DMZ)

<https://kb.rice.edu/108626>

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

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

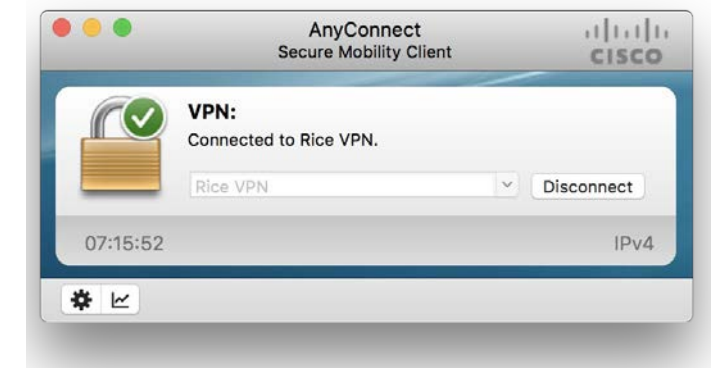
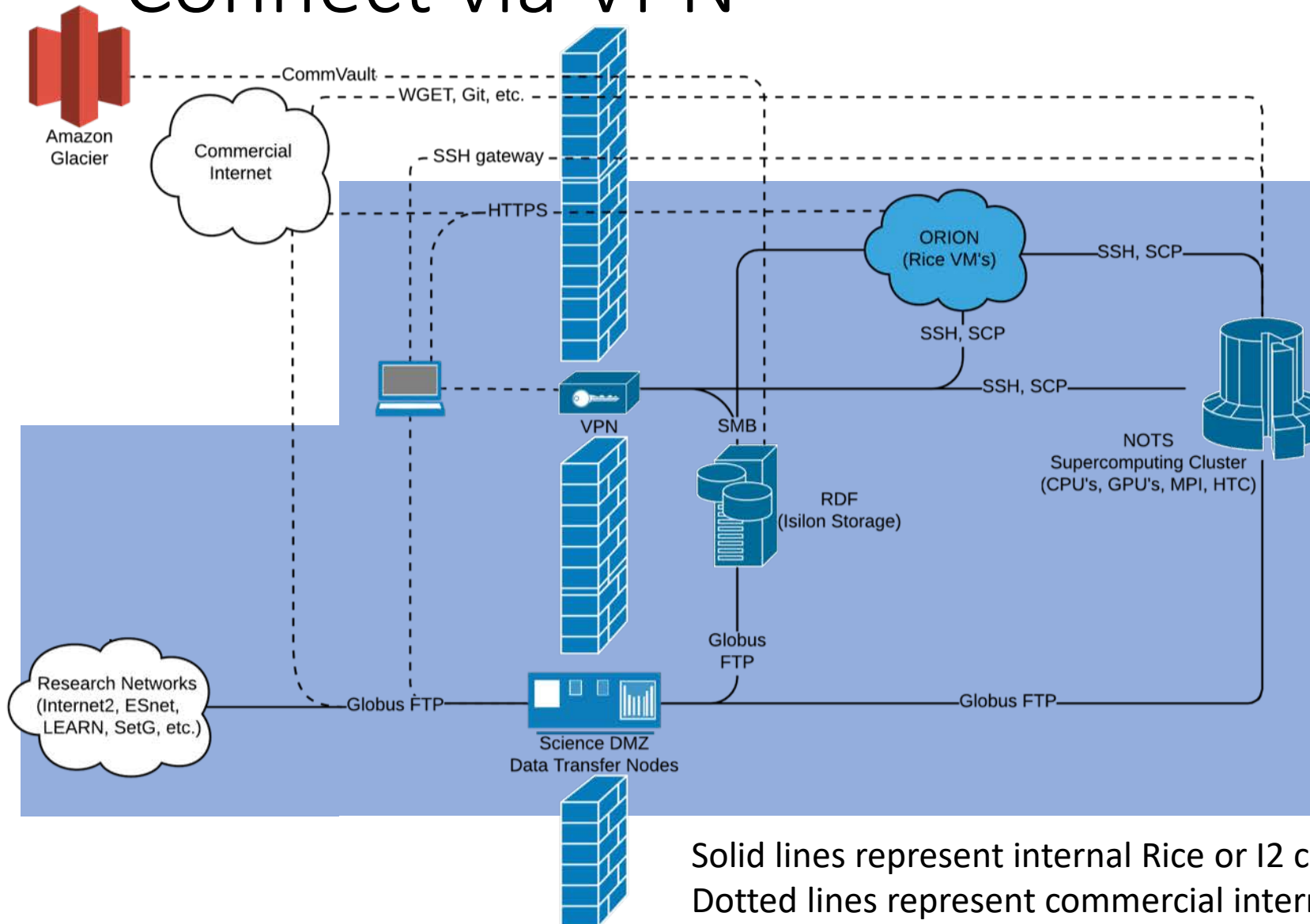
<https://kb.rice.edu/108241>

RDF EXERCISE

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

VPN Examples

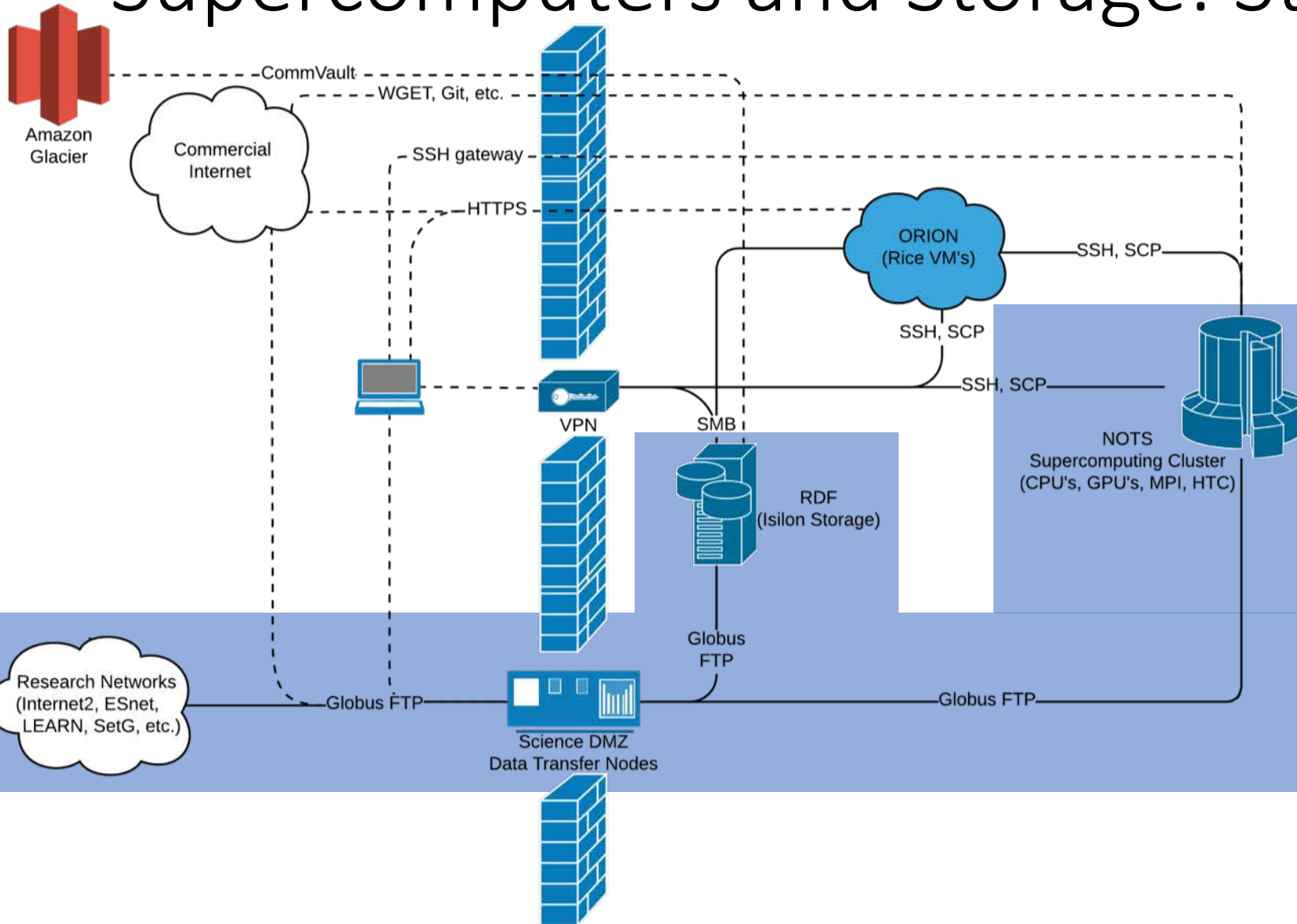
Connect via VPN



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

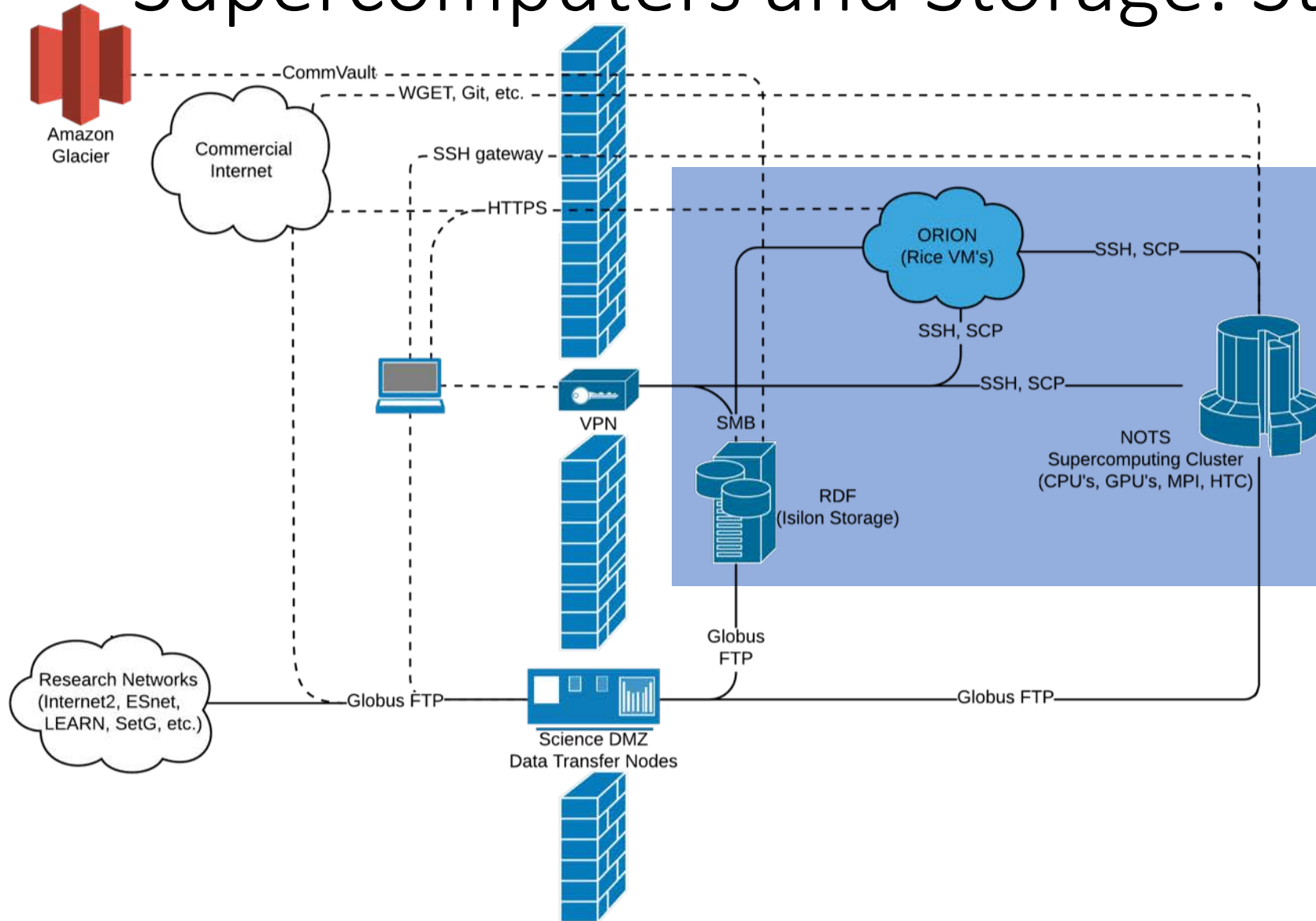
For medium datasets, you can:

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

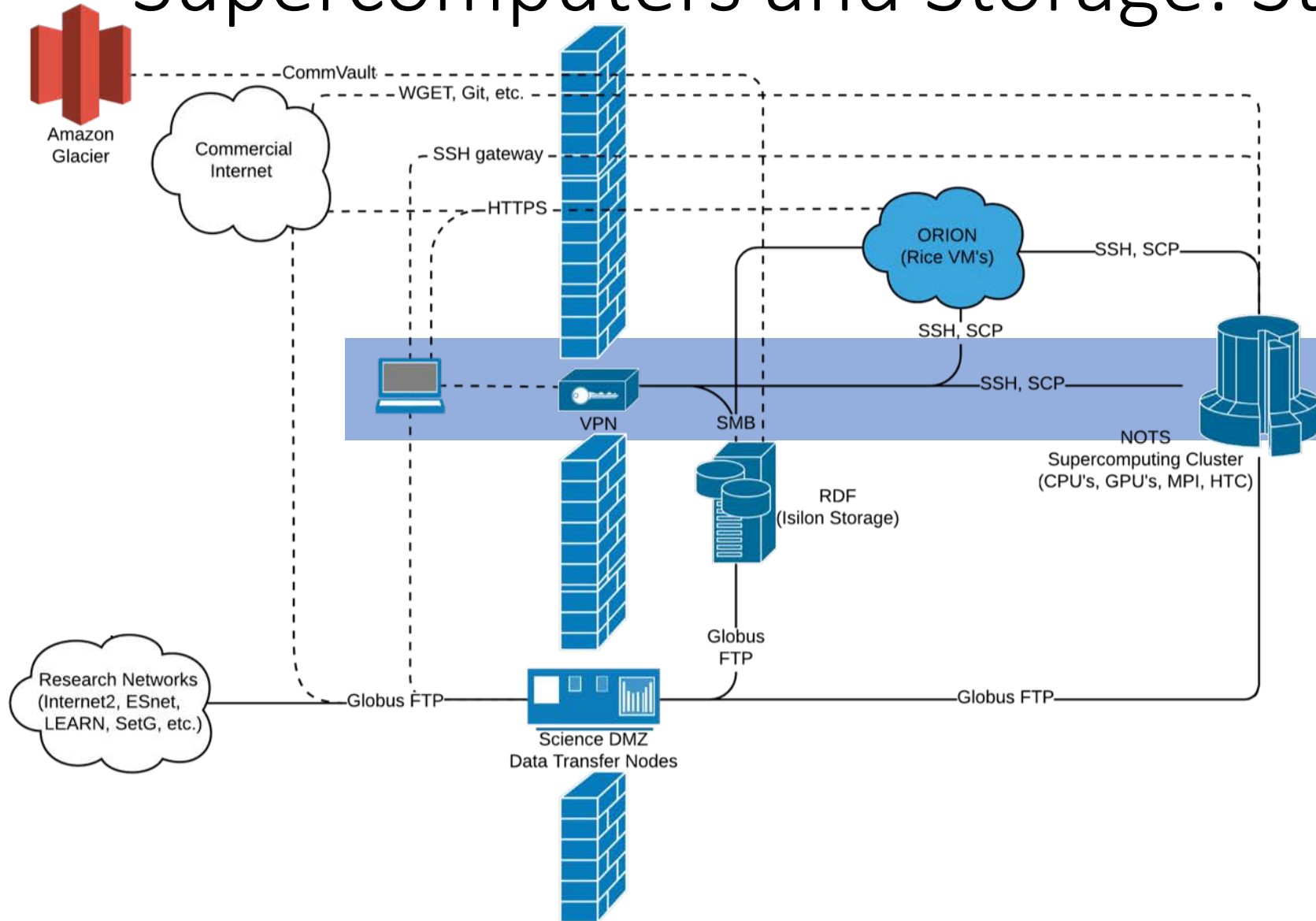
For medium datasets, you can:

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

The screenshot displays the Globus File Manager web interface. The browser address bar shows the URL: https://app.globus.org/file-manager?destination_id=5416d524-fb41-11e9-8a5d-0e35e6625. The interface is split into two panels. The left panel shows the source endpoint: 'Collection: NOTS scratch' and 'Path: /jcm10/'. The right panel shows the destination endpoint: 'Rice University CRC RDF endpoint' and 'Path: /rdf/jcm10/riedeltest/'. A central menu is open, listing actions such as 'Share', 'Transfer or Sync to...', 'New Folder', 'Rename', 'Delete Selected', 'Download', 'Open', 'Upload', 'Get Link', 'Show Hidden Items', and 'Manage Activation'. The file list on the right shows a table with columns for NAME, LAST MODIFIED, and SIZE. The file 'tilesout-march27' is highlighted in blue. At the bottom, there are buttons for 'Start' and 'Transfer & Sync Options'.

| NAME | LAST MODIFIED | SIZE |
|--------------------|-------------------|----------|
| riedel_jan7.tar.gz | 01/07/2020 10:... | 19.06 MB |
| run notes.txt | 12/02/2019 02:... | 1.29 KB |
| SanbornBlocks | 11/06/2019 04:... | - |
| SanbornLots | 11/06/2019 04:... | - |
| Tiles | 11/20/2019 10:... | - |
| tilesout-march25 | 03/26/2020 12:... | - |
| tilesout-march26 | 03/26/2020 09:... | - |
| tilesout-march26a | 03/26/2020 06:... | - |
| tilesout-march27 | 03/27/2020 02:... | - |

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

The screenshot shows a web browser window with the URL <https://kb.rice.edu/108242>. The page is titled "CRC Getting Started on Globus DTN network" and includes an introduction and a list of DTN nodes. The introduction states: "Rice University CRC's Globus Data Transfer Node (DTN) network nodes provide massively parallel file transfers to facilitate moving large data sets efficiently. A transfer to similarly configured Endpoints will take advantage of high-throughput networking. You will be able to move data between participating institutions via Globus, including between our DTN endpoints. Additionally, if you install and configure the Globus Connect software you will be able to transfer files using your local computer as an Endpoint." The DTN nodes list includes:

- Rice University CRC DTN for /scratch on NOTS
 - facilitates the movement of data into and out of the "/scratch" shared filesystem available for NOTS.
- Rice University CRC DTN for /home on NOTS
 - facilitates the movement of data into and out of the "/home" shared filesystem available for NOTS.
- Rice University CRC DTN for /projects on NOTS
 - facilitates the movement of data into and out of the "/projects" shared filesystem available for NOTS.
- Rice University CRC deepsc in storage condo
 - facilitates the movement of data into and out of the "/storage/condo/" shared filesystem available for NOTS.
- Rice University CRC Data transfer node Google Drive connector for "Shared Drives"
 - provides access to create a collection using your Rice Google account ("Shared Drives") as an endpoint allowing you to transfer data between your Rice Google Drive account and other Globus endpoints.
- Rice University CRC Data transfer node for Google Drive connector for "My Drive"
 - provides access to create a collection using your Rice Google Drive account ("My Drive") as an endpoint allowing you to transfer data between your Rice Google Drive account and other Globus endpoints.

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

DATA xFER EXERCISE

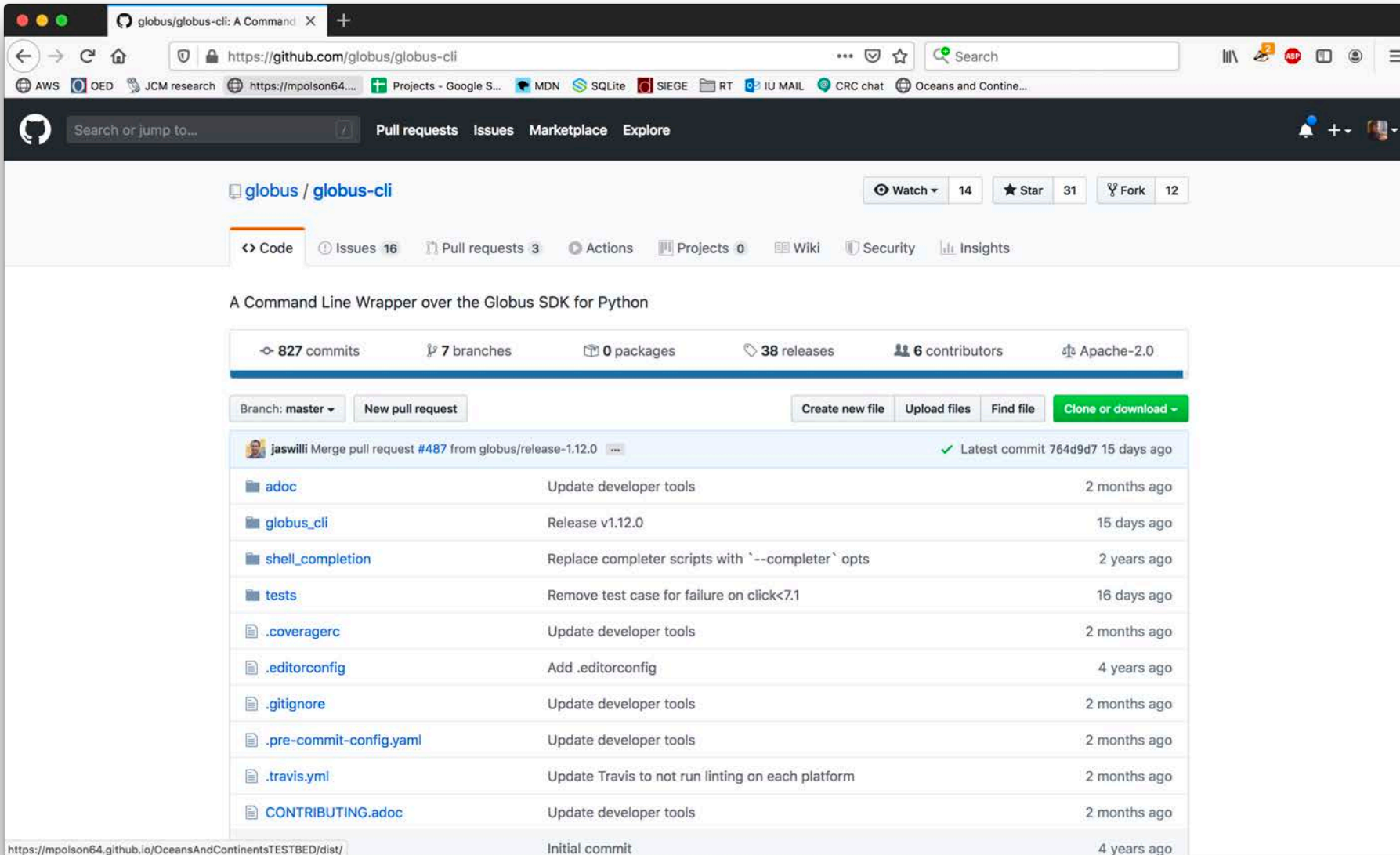
- GLOBUS

- 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



globus / globus-cli

Watch 14 Star 31 Fork 12

Code Issues 16 Pull requests 3 Actions Projects 0 Wiki Security Insights

A Command Line Wrapper over the Globus SDK for Python

827 commits 7 branches 0 packages 38 releases 6 contributors Apache-2.0

Branch: master New pull request Create new file Upload files Find file Clone or download

| Commit | Description | Time |
|--|---|--------------|
| jaswilli Merge pull request #487 from 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 scripts with `--completer` opts | 2 years ago |
| tests | Remove test case for failure 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 linting on each platform | 2 months ago |
| CONTRIBUTING.adoc | Update developer tools | 2 months ago |
| https://mpolson64.github.io/OceansAndContinentsTESTBED/dist/ | Initial commit | 4 years ago |

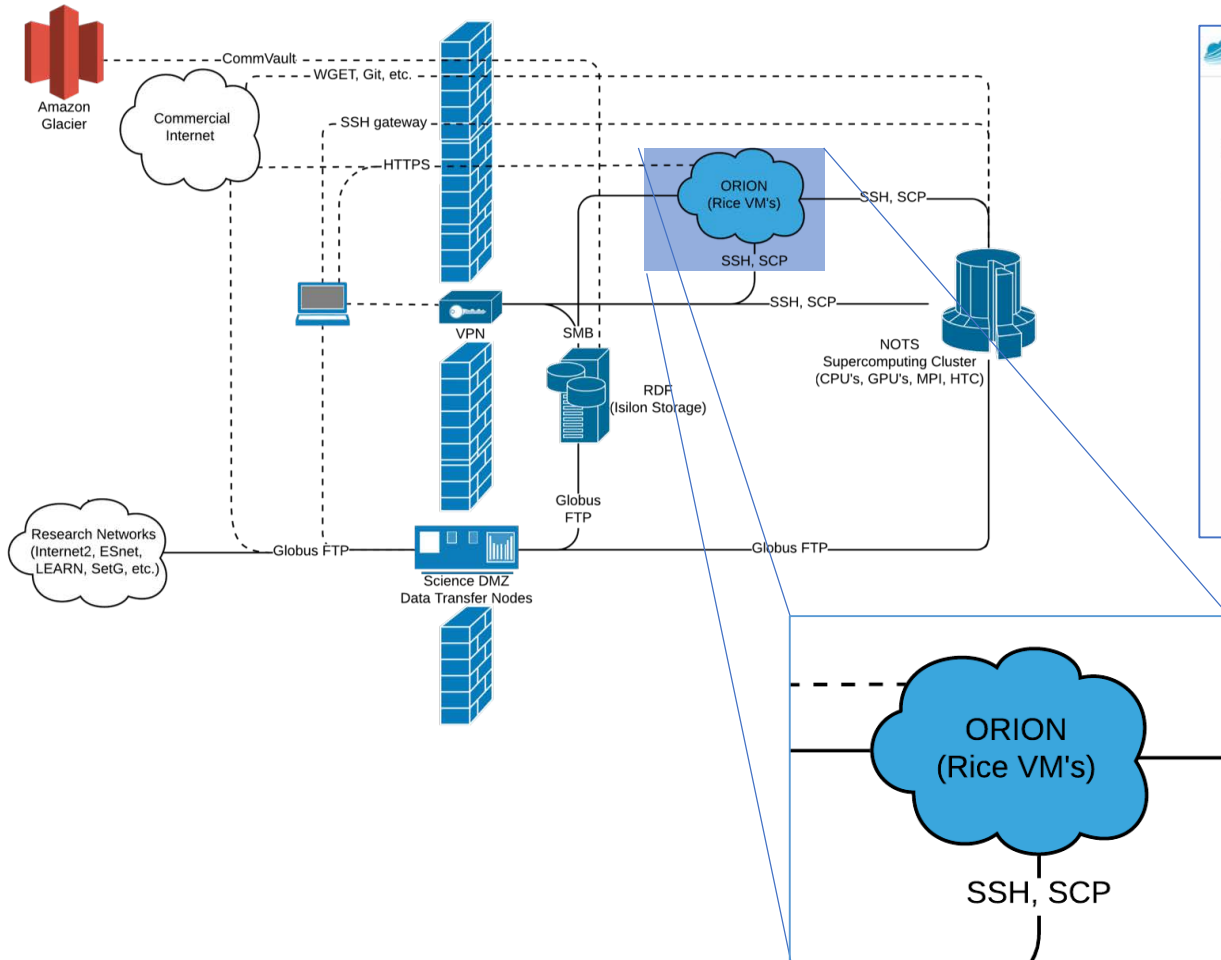
Using Globus' CLI to automate tasks

- 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)

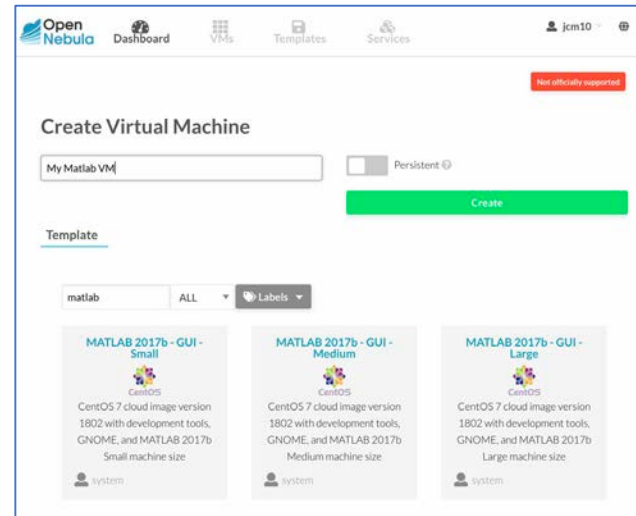
<https://github.com/globus>

ORION Virtual Machines

ORION Virtual Machines

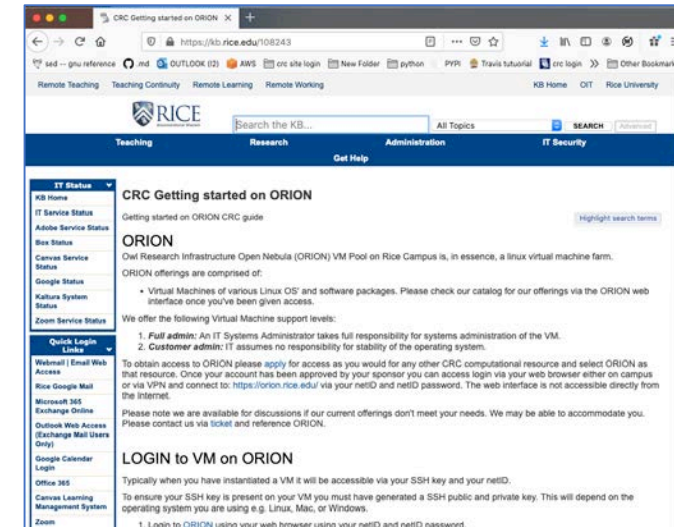


Web interface to manage VM's



<https://orion.crc.rice.edu>

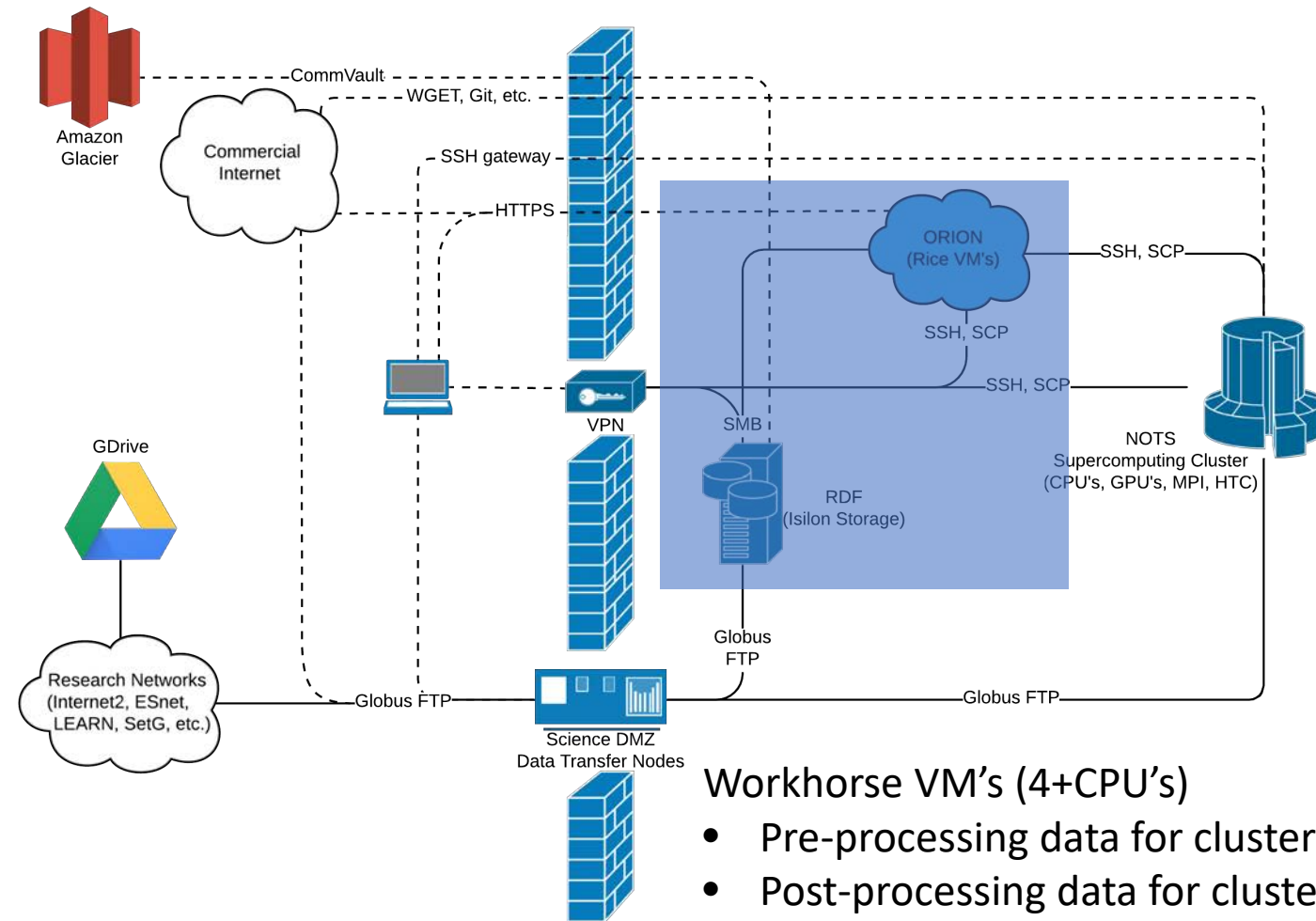
Orion Documentation



<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



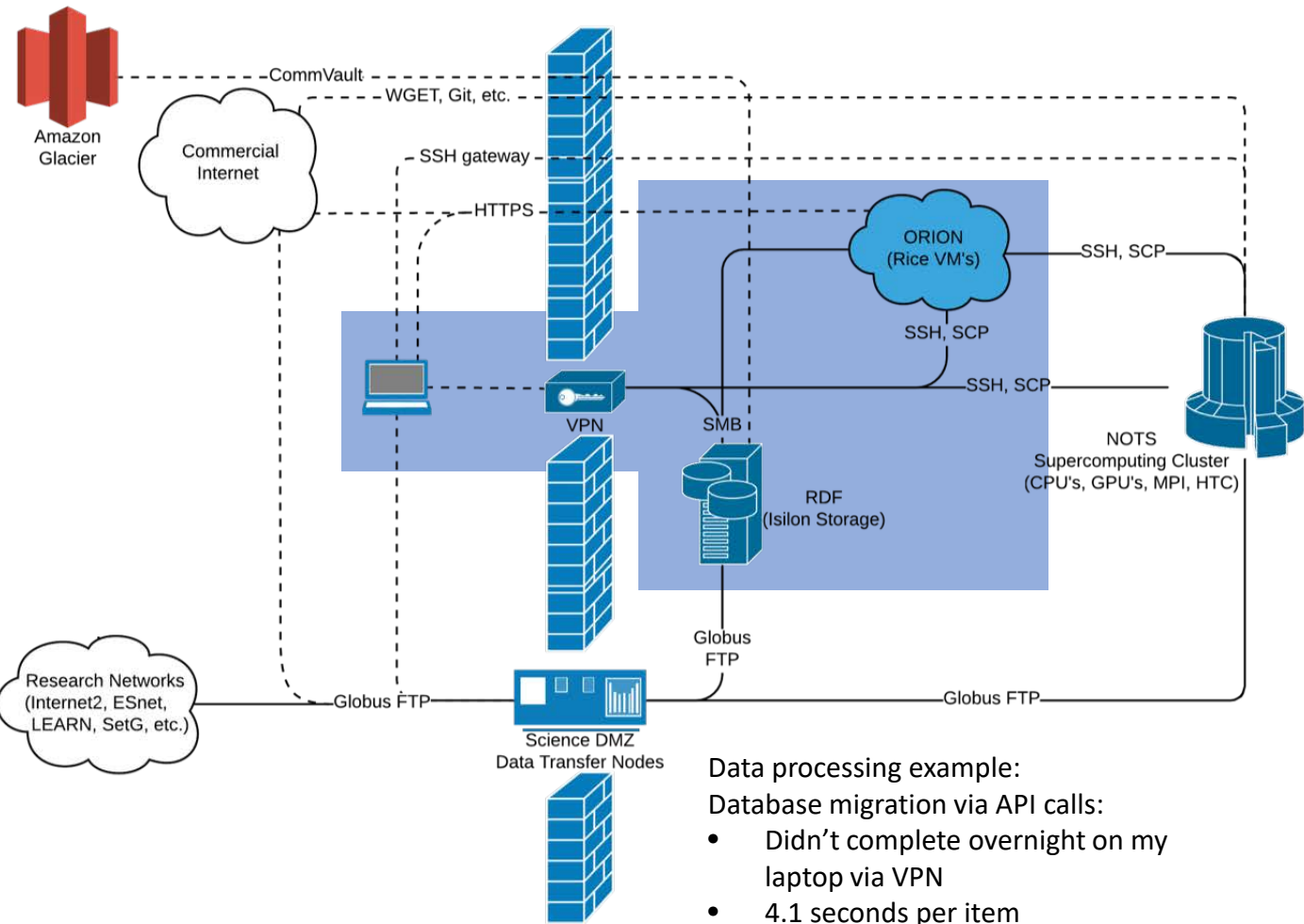
Workhorse VM's (4+CPU's)

- Pre-processing data for clusters
- Post-processing data for clusters
- Medium-sized jobs
- Interactive jobs

Performance: Lookups per second vs. processes count, finding 1000 matches

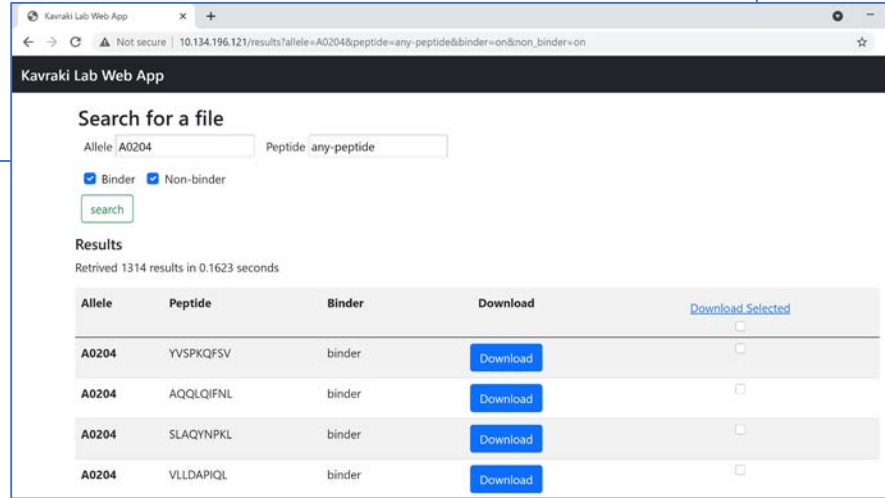
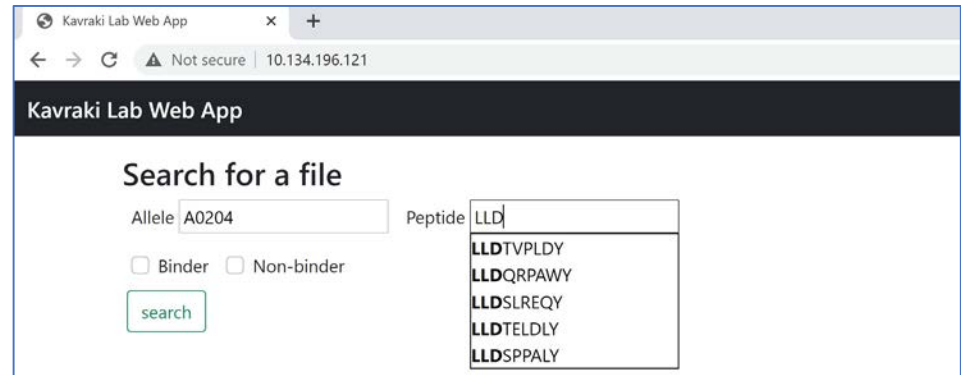


Lightweight VM with Mounted Storage, Private IP



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



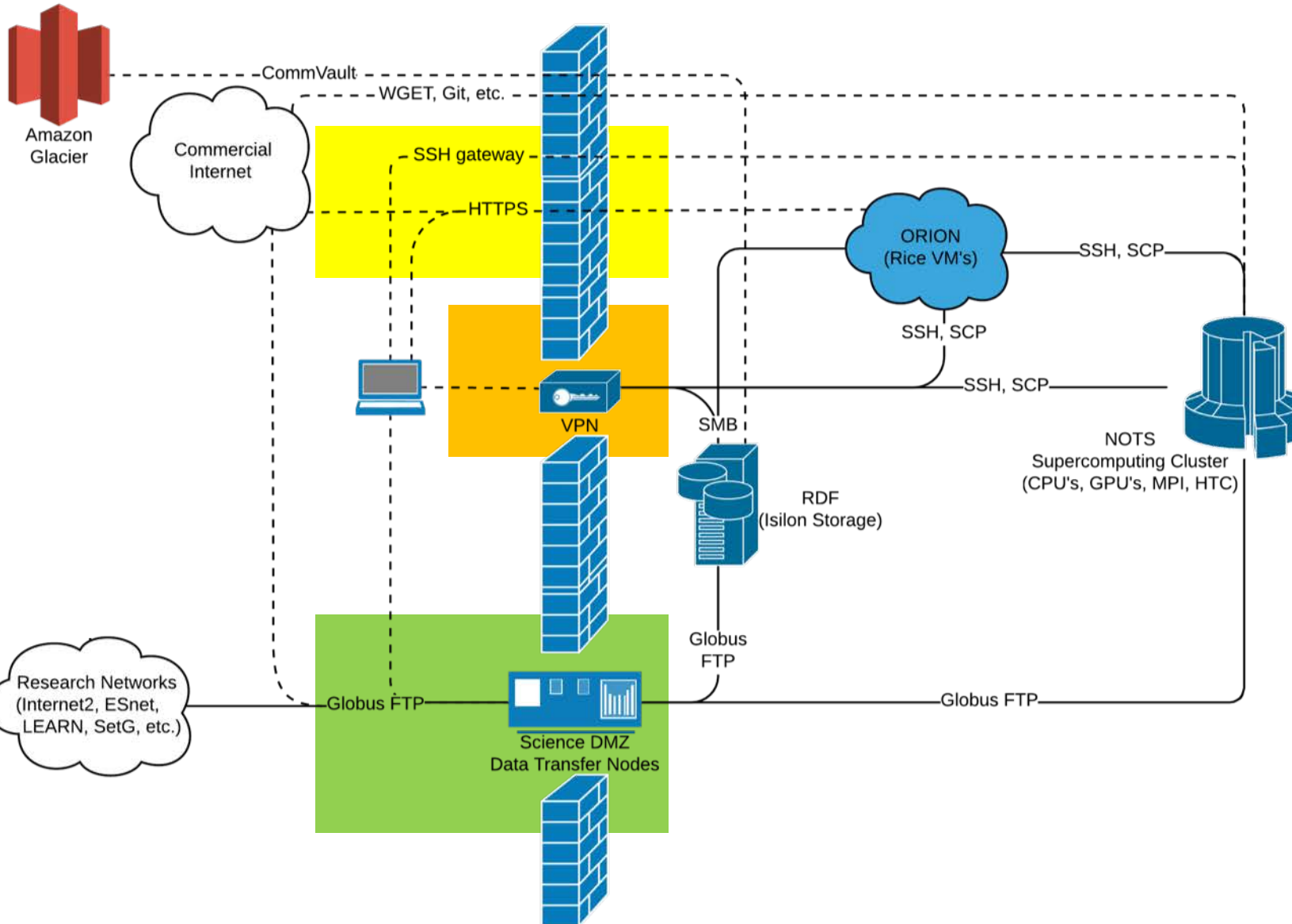
Data Interface Example:
 Lightweight Flask app connecting to RDF datasets
 (courtesy of Kavradilab: Anja Conev, Nonso Chukwurah, Romanos Fasoulis)

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
- helpdesk@rice.edu
- Facilitators:
 - Clinton Heider: heider@rice.edu
 - John Mulligan: jcm10@rice.edu