Virtual Research Desktop Environment (VRDE)

A computing environment designed for research involving regulated data

Clinton Heider, Center for Research Computing
Hien Le, Kinder Institute for Urban Research

What is VRDE?

- CRC managed computing environment designed to comply with NIST 800-171 security controls
- VRDE is for RESEARCH ONLY. It is not approved to house any data related to university administration, operations, departmental data, student records, etc.
- Uses VDI (Virtual Desktop Interface) technology to allow access to a computing platform which is otherwise isolated on the network
- Available to Rice researchers who need to work with regulated data, specifically CUI (Controlled, Unclassified Information)
- VRDE was originally built for the Kinder Institute, which is the largest user. However, the system is designed to house multiple tenants, each of which has complete network isolation from other tenants. Each research group can be assured that their data and compute systems are discrete from other tenants.

CUI – Controlled, Unclassified Information

CUI is non-classified information (i.e. information the Government creates or possesses, or that an entity creates or possesses for or on behalf of the government) that requires safeguarding or dissemination controls compliant with law, regulations, and government-wide policies.

May include items such as:

- **Export controlled technology and information**
- **Proprietary business information**
- **Federal statistical data such as census data**
- **Information protected by HIPAA and FERPA**

Research Compliance

CUI

https://research.rice.edu/compliance/cui

Export Controlled Technology

https://research.rice.edu/compliance/export-control-overview

NIST 800-171

A set of controls and requirements for non-Federal information systems which store, process or transmit CUI

- Role-based access controls, identity management and authentication
- Audit & accountability
- Information integrity & protection (physical media, facilities, data transmission, backups)
- Systems security, configuration management, maintenance
- Risk assessments – initial and periodic
- Incident Response
- Documentation of procedures
- User training
- Disaster Recovery Plan

Roles and Responsibilities

SPARC (Sponsored Research & Research Compliance)

- Manages the University's data compliance responsibilities
- Assists researchers with developing proposals, approves data use agreements and technology control plans

OTT (Office of Technology Transfer)

- Works with the Office of the General Counsel to draft & review data use agreements, MOUs, etc.

ISO (Information Security Office)

- Performs risk assessments for projects involving controlled data.
- Evaluates and recommends required technology controls
- Provides security monitoring tools and log reports to VRDE tenants and OIT staff.
- Conducts audits of systems which house regulated data.

Roles and Responsibilities (cont.)

CRC (Center for Research Computing)

- Provides secure infrastructure for data storage & analysis which meets or exceeds 800-171 standards
- Manages access & network controls, software deployment, patches and security updates
- Maintains documentation of the secure infrastructure and operational procedures
- Consults with researchers in evaluating computing requirements for research projects

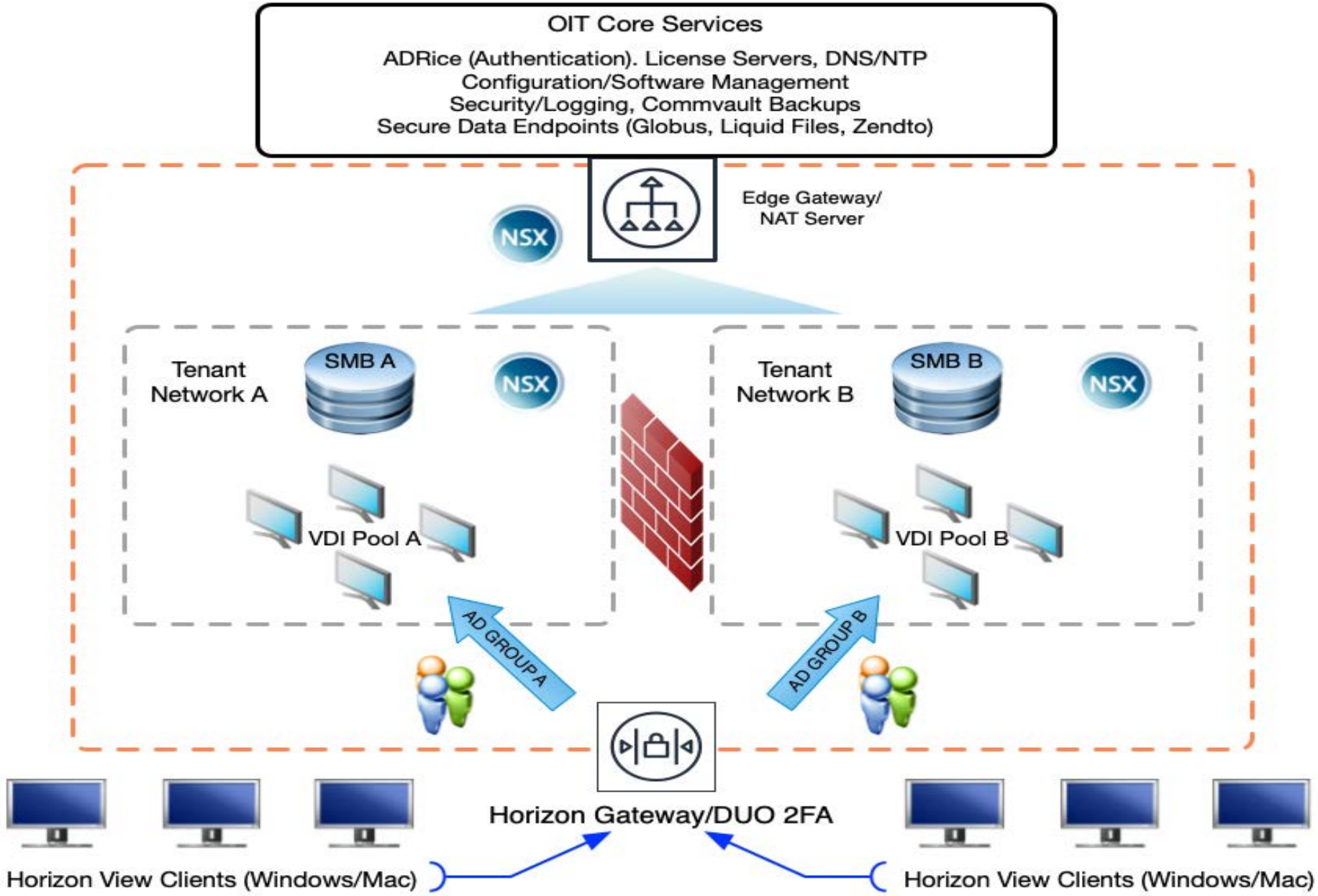
Tenant (Research Group)

- The researcher is the subject matter expert and will need to know which datasets are controlled, who may access them, and whether data are properly de-identified
- Data management is the responsiblity of the researcher. This includes periodically reviewing access controls and documenting procedures for handling, managing, processing or analyzing controlled data. Smaller tenants which are dealing with few projects and a small number of users may handle these tasks themselves, but larger centers and research groups may want dedicated data managers for this purpose.
- CRC/ISO may provide tools or reports to the tenant to support these processes, but the processes themselves are the responsiblity of the researcher.

VRDE Security Architecture

- VMWare VDI (Virtual Desktop Interface) instances are accessed using Horizon View client with DUO two factor authentication. Data transfer to and from VDI instances via USB, local drives, copy/paste, printing, etc. is blocked; all the user gets are pixels. VDI systems are blocked from accessing external networks or other internal tenant networks. They can only reach designated core OIT services (authentication, licensing, DNS, approved data transfer points, monitoring & security systems, etc.)
- Tenant data is stored on an internally addressed network storage appliance (Dell Unity) which provides a Windows (SMB) share to each tenant's private, isolated virtual network. Data is encrypted at rest. SMB traffic never leaves the tenant's private network.
- Rice's Active Directory system authenticates users and assigns roles based on membership in Active Directory groups. These roles can in turn be used to access the correct VDI pool and grant or restrict access to specific projects/data sets using NTFS file permissions
- Authentication tied to NetID ensures compliance with many core 800-171 standards including password strength, lockout controls, prevention of shared credentials, accountability, auditing & logging and institutional validation of authorized users

VRDE Security Architecture



VRDE Security Architecture (continued)

- CRC performs periodic maintenance to ensure that all system components have security updates. System events and user logins are logged to an external log system (Splunk). Tenants can request Splunk reports from Rice ISO. Reviews of, or changes to, Active Directory access control groups can be requested with help desk tickets routed to the CRC. Only authorized CRC/OIT personnel have administrative access to VRDE components.
- VRDE Hardware is on-premise, in Rice's Primary Data Center (PDC), which can be accessed only by authorized personnel.
- The PDC has redundant power and networking capabilities which help to ensure stability and uptime during maintenance or power interruptions. PDC staff additionally have documented data destruction procedures for the disposal of equipment which may contain CUI.
- Backups of data may be requested by the tenant, but are optional. Backups are made directly from the storage appliance using OIT's Commvault software. Archives are encrypted and stored in the AWS Glacier cloud storage service.

VRDE Data Transfer

- VDI instances have no internet access or direct access to campus storage
- Data transfers must be through identified, audited, managed endpoints over encrypted channels: some examples
  - Liquid Files – Commercial tool used extensively by Kinder. Good for multi user, multi-project, frequent transfers which require an audit trail. License required.
  - ZendTo (https://zendto.rice.edu/) – email with URL, encrypted transfers, option to encrypt files. Limited transfer size. More details at Knowledge base article 83248
  - GLOBUS FTP – federated file transfer network. Supports encryption, good for large transfers globus.org
  - scp/sftp from secured endpoints. Manual audit of transfers (maintain a log)
- ***Arbitrary access to external software repositories is not possible*** - i.e., python pip repositories, github, R cran. Tenants are typically responsible for developing and maintaining their own software stack. This process can be challenging, so plan accordingly - where possible, test your app stack thoroughly on conventional systems with safe data, then work with CRC to move that stack into the VRDE environnment.

VRDE VDI Specs/Cost

Typical configurations

VDI Class	Cores	RAM (GB)	GPU RAM (GB)	Cost/Yr
Small	2	8	0	$275.41
Medium	4	16	1	$365.89
Large	16	64	2	$799.96

Data Storage cost

Storage Cost/TB	Cost/Yr
Flash	$1092.91
Conventional Disk	$98.00

GPU capability/licensing

NVIDIA GPUs require a license to have resources shared among VDI instances, this cost is included in the GPU capable VDIs listed above. Note that existing GPU capability is intended to support graphical applications requiring GPUs (ArcGIS Pro, etc.) and are not recommended for machine learning/GPU compute tasks

VRDE

An End-User and Data Manager Perspective

Hien Le
Director of Data Operations, Kinder Institute for Urban Research


AIRPORT

Types of Data Enclaves

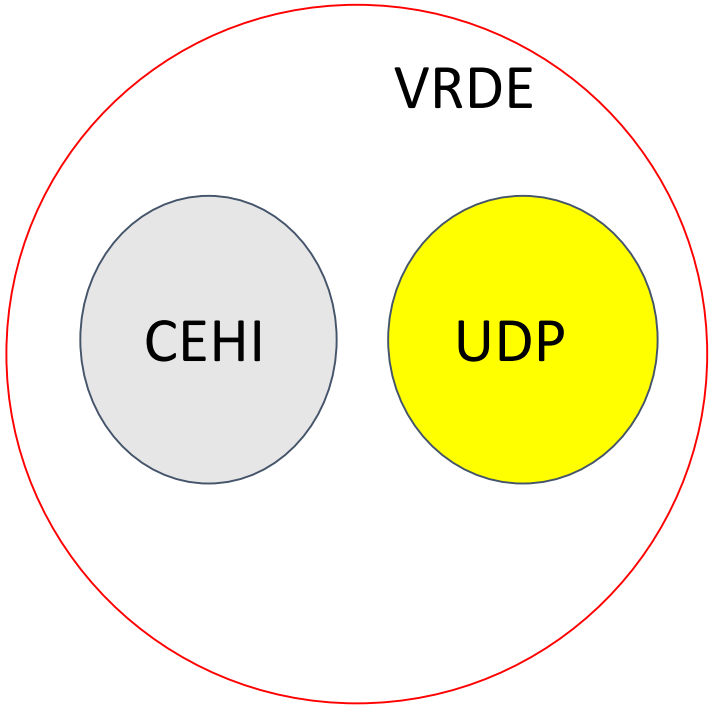
- Brick and mortar
- On-premise
- Cloud

VRDE User Experience

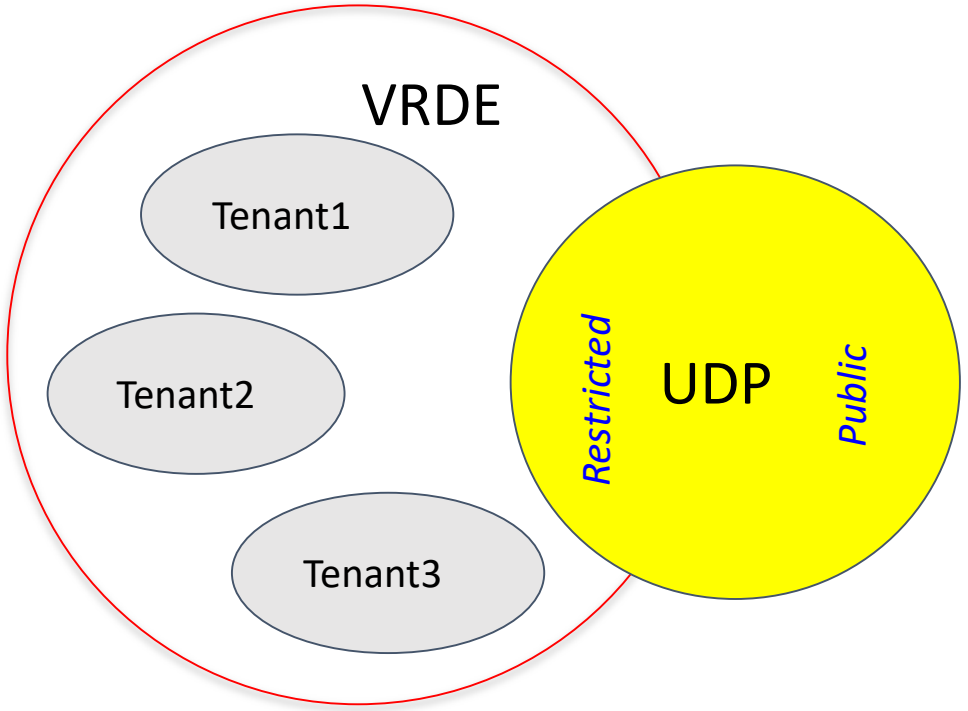
- User: researchers vs. (data) managers
- Security spectrum: Managing expectations
- Scalability of the solution
- Accessing a Windows 10 VM without access to most Internet resources. This implies restrictions on:
  - Applications
  - File transfers

VRDE and Kinder Institute's UDP

UDP = Urban Data Platform, a data repository originally founded by Dr. Katherine Ensor, hosted at the Kinder Institute for Urban Research



In the beginning



Now

Software and development environments within VRDE

- Licensing
  - Site license
    - Applied once vs. contacting the site licensing server
  - Non-site license: What happens if there are less licenses than VMs? Sassafras Keysight
- Development environments with extensions or packages
  - R
    - CRAN
    - Other repositories / Special cases
  - STATA
    - SSC
    - STATA journal
  - Python
  - SAS
- Dashboard?

Data Access Within VRDE

- Public vs. Restricted
- IRB Protocol
- Overarching/Global IRB Protocol for Data Managers
- Non-IRB requirements (DSA/DUA/License-specific requirements)
  - Rice vs. non-Rice
  - List of users
  - Deletion
  - Annual reports
  - Audits

Data Access/Organization Within VRDE

- Folder-base system
- Use of the (My) Documents folder
- Duplication of datasets
  - Library folders / Project folders
- AD groups
- Data management and user management
- User training

File Transfers

- What are your requirements?
  - What can be transferred?
    - Inbound
      - Audit requirements
      - Archiving/preservation requirements
    - Outbound
  - Logging requirements
  - Microdata vs. aggregated data
    - Microdata: de-identification
    - Aggregated Data: suppression of small values, limit the level of n-way freqs (cross-tabulations)
  - Who can transfer files?
- LiquidFiles

Issues/Challenges

- Logout limbo
- Roaming profile not loading
- VM hugging
- Review of files to export
- Working on resource-hungry projects
  - Parallel-processing
    - Explicit code
    - Application limits
  - Temporary files / Work folder

VRDE Demonstration

VRDE access requires the VMWare Horizon View client. This software is available for both PC and Mac and can be downloaded from the VMWare website directly, or internally from the Rice Knowledge Base page which includes download links and instructions:

Mac instructions: Knowledge base article 114679

Windows instructions: Knowledge base article 115348

VRDE utilizes DUO two factor authentication and a valid Rice NetID is required! You do not have to be connected to the Rice VPN or on Rice campus to reach the connection gateway.

Demonstration:

- Launch the Horizon View client and designate the VRDE Horizon gateway hostname (vrde.crc.rice.edu)
- Authenticate using your NetID and NetID password and DUO 2FA
- Select the appropriate VDI pool (most users will only see one)
- The user is already authenticated by Horizon and the session will login automatically
- Network drives and profiles will be automatically assigned at login
- If you have an existing login on a VDI, it will reconnect you to that login session. Likewise, if you disconnect without logging out, your session will continue (including running jobs). If you logout, your session will terminate and at next login, you will be assigned a new VDI instance.