

Regional Data DMZ in Support of Distributed Applications in Pennsylvania and New Jersey

**Internet2 Technology Exchange
October 17, 2017**

**Gwendolyn Huntoon - KINBER
Kenneth Miller - Penn State
Jim Stankiewicz - NJEDge.Net
Barr von Oehsen - Rutgers**

Agenda

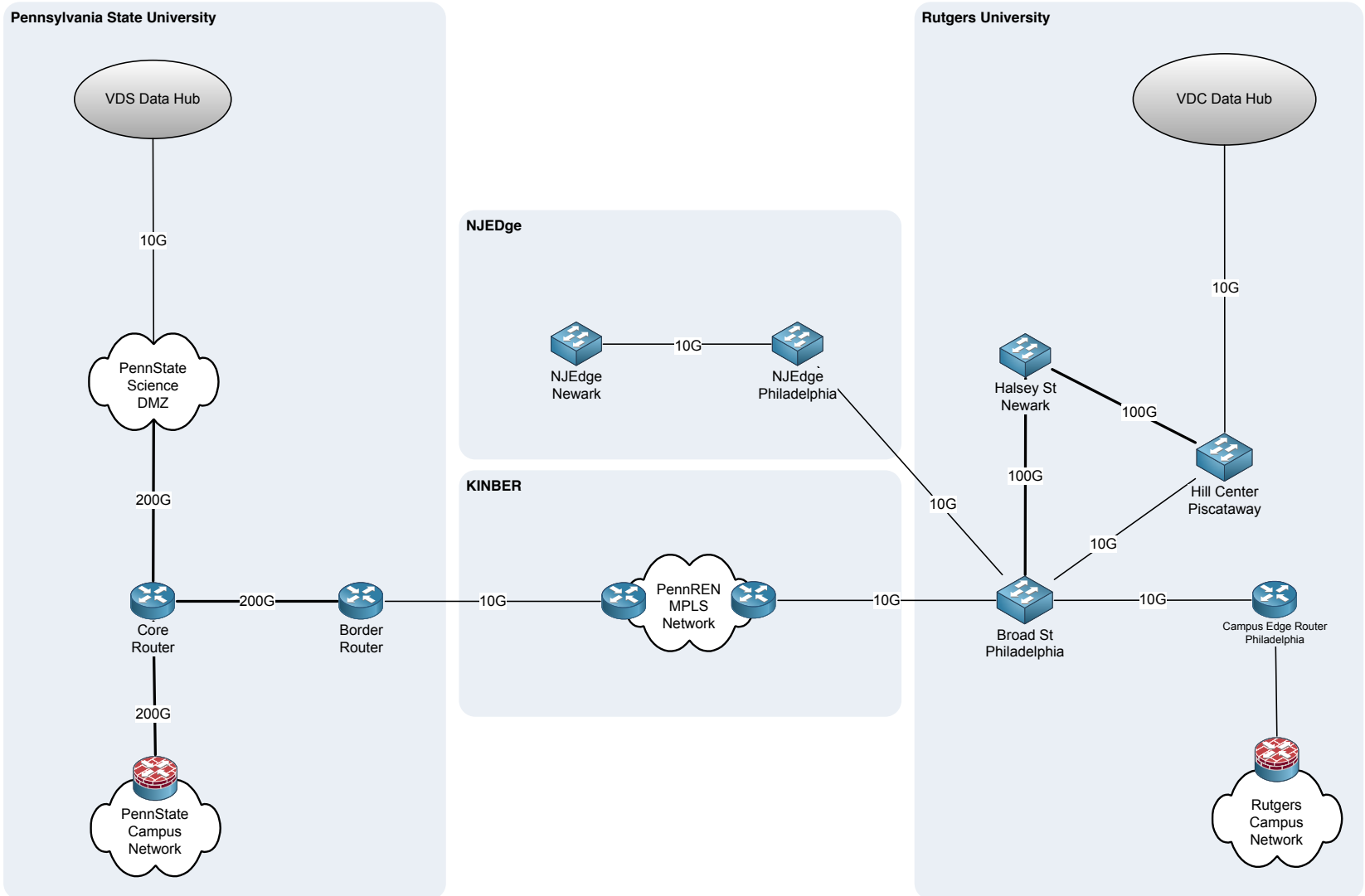
- Motivation
- Campus Network
- Regional Network
- Current Status
- Next Steps

Regional Data DMZ

- High-bandwidth network connectivity between the regional networks, PennREN (KINBER) and NJEdge as well as individual organizations located in Pennsylvania and New Jersey.
- Long-term goal of connecting this regional infrastructure to the broader community via regional and national network connections, including Internet2.
- Initial application to utilize this infrastructure is the Virtual Data Collaboratory (VDC).
 - NSF funded DIBBS project with the goal of establishing state of the art data-intensive computing solutions, storage, and networking, federated with an innovative data services layer

Data DMZ Topology

VDC Data DMZ Topology

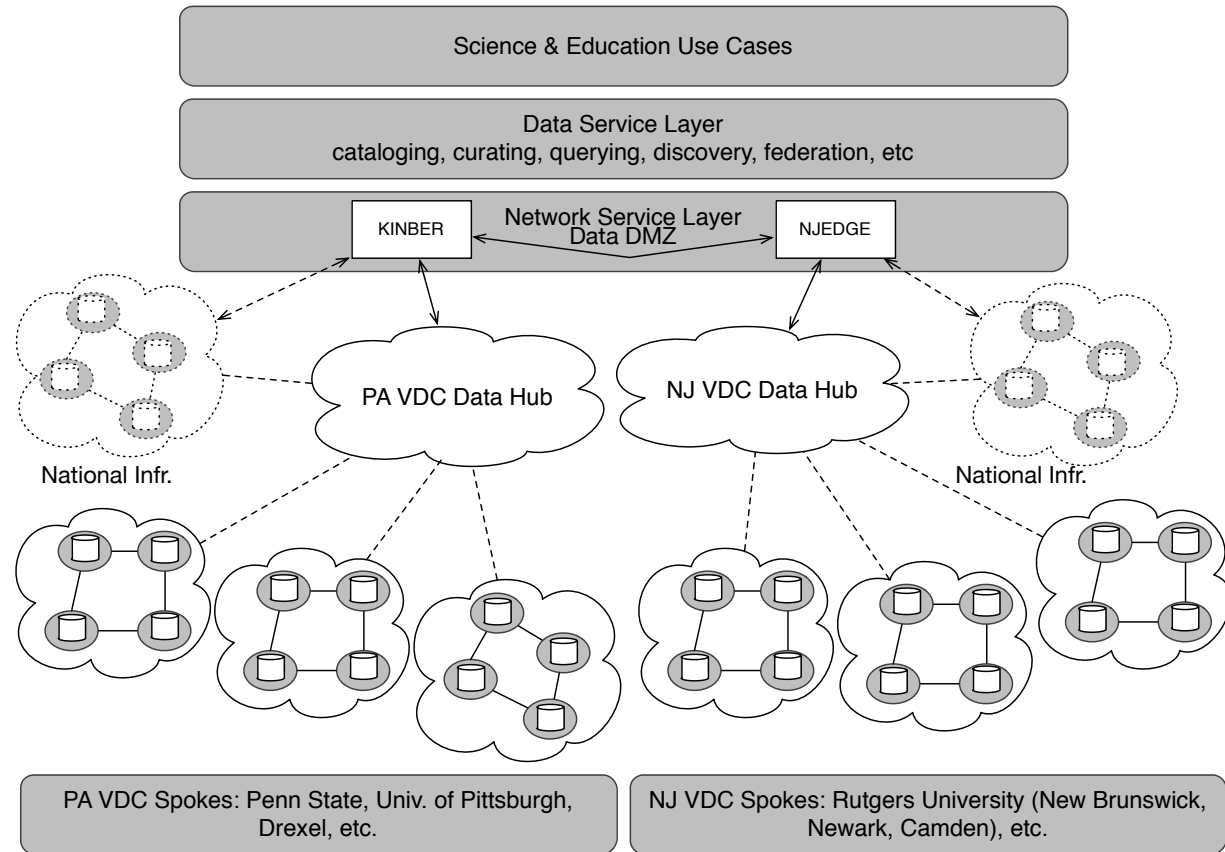


VDC Application

- CIF21 DIBBs: EI: Virtual Data Collaboratory: A Regional Cyberinfrastructure for Collaborative Data Intensive Science (NSF Award 1640834)
 - PI: Manish Parashar; Co-PI: Grace Agnew, James von Oehsen, Jenni-Louise Evans, Vasant Honavar
 - Collaboration between Rutgers, Penn State, NJEdge and KINBER
 - Leverages concepts from the PRP
- Applications
 - Deciphering Sequence and Structural Correlates of Protein Nucleic Acid Interactions (H. Berman, RU & V. Honavar, PSU)
 - High-Volume City Data Sharing and Processing for Smart, Resilient, and Sustainable Cities (J. Gong, RU; Z. Zhu, CUNY; X. Liang, University of Pittsburgh; M. Balduccini, Drexel University)
 - Ocean Observatories Initiative (I. Rodero, RU, M. Parashar, RU)

VDC Architecture

- VDC Hub: Scalable data-intensive computing platform
- Data services to support research workflows
- Regional science data DMZ network



VDC Data DMZ

- Supports data import/export services with necessary qualities of services to enable efficient and transparent access to data and compute regardless of a scientist's location.
- Components
 - Data DMZ backbone consists of direct 10 GE connections between Rutgers, NJEdge and KINBER
 - Data Hubs connect directly to the Data DMZ or through their regional network
 - Data Spokes connect through the regional network
 - FIONAs deployed at Hub and Spoke sites

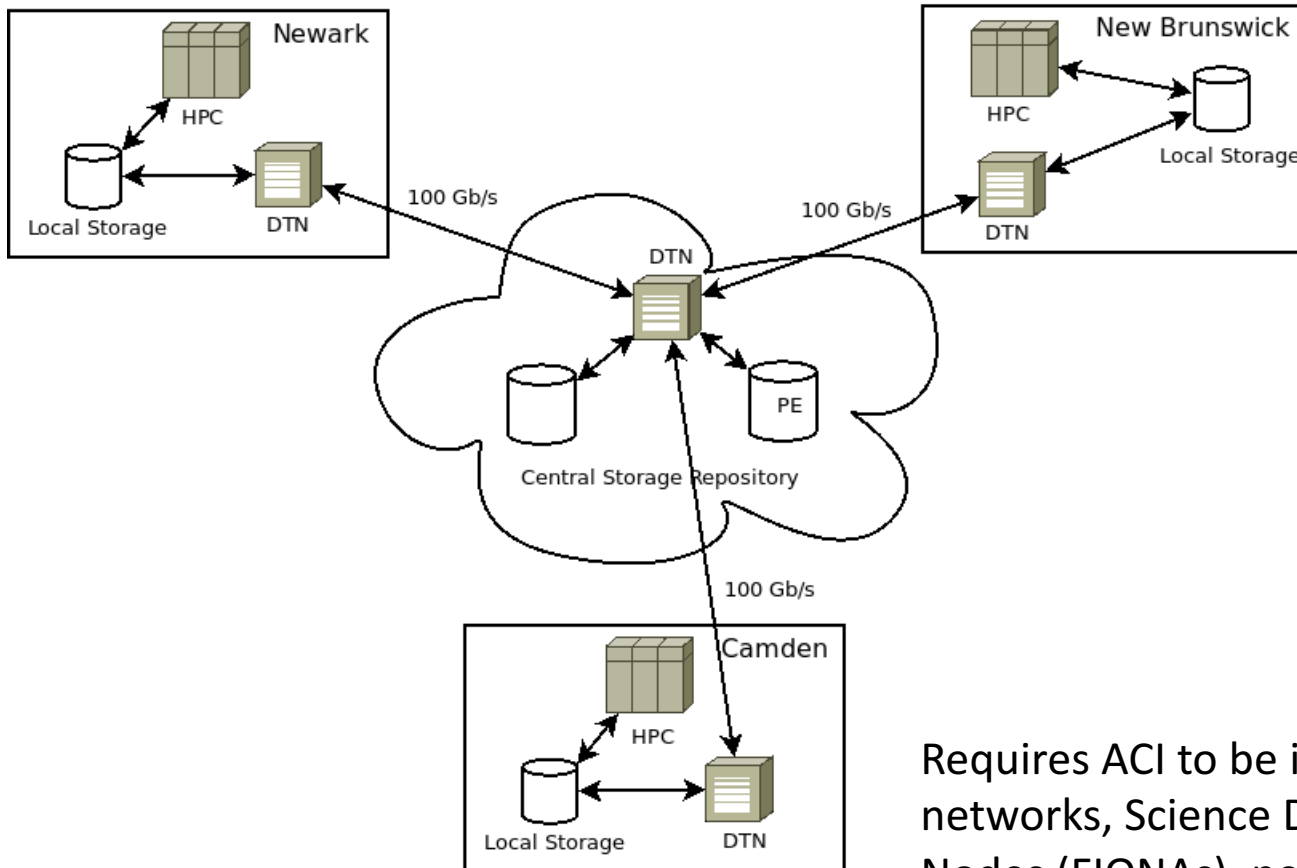
VDC Project Components

- VDC Hubs – system sites
 - Rutgers University – Main Campus
 - Penn State – Main Campus
- VDC Spokes – collaboration sites
 - New Jersey – Rutgers Camden, Rutgers Newark
 - Pennsylvania – Drexel, Temple, University of Pittsburgh
 - New York – City University of New York
- Data DMZ Backbone
 - NJEdge and PennREN (KINBER)

Rutgers/New Jersey Applications

- Brain Research
- Medical Imaging
- Genomics
- Genetics
- Bioinformatics
- Observational Astronomy
- Experimental Physics
- Climate Modeling
- Network Research Communities to fully utilize CICNet

Rutgers DMZ – Local Research Platform



Requires ACI to be in place – fast/low-latency networks, Science DMZs, SDN, Data Transfer Nodes (FIONAs), perfSONAR, XDMoD, Advanced Computing, and Storage. Building block for region, national...

One Rutgers Research Platform

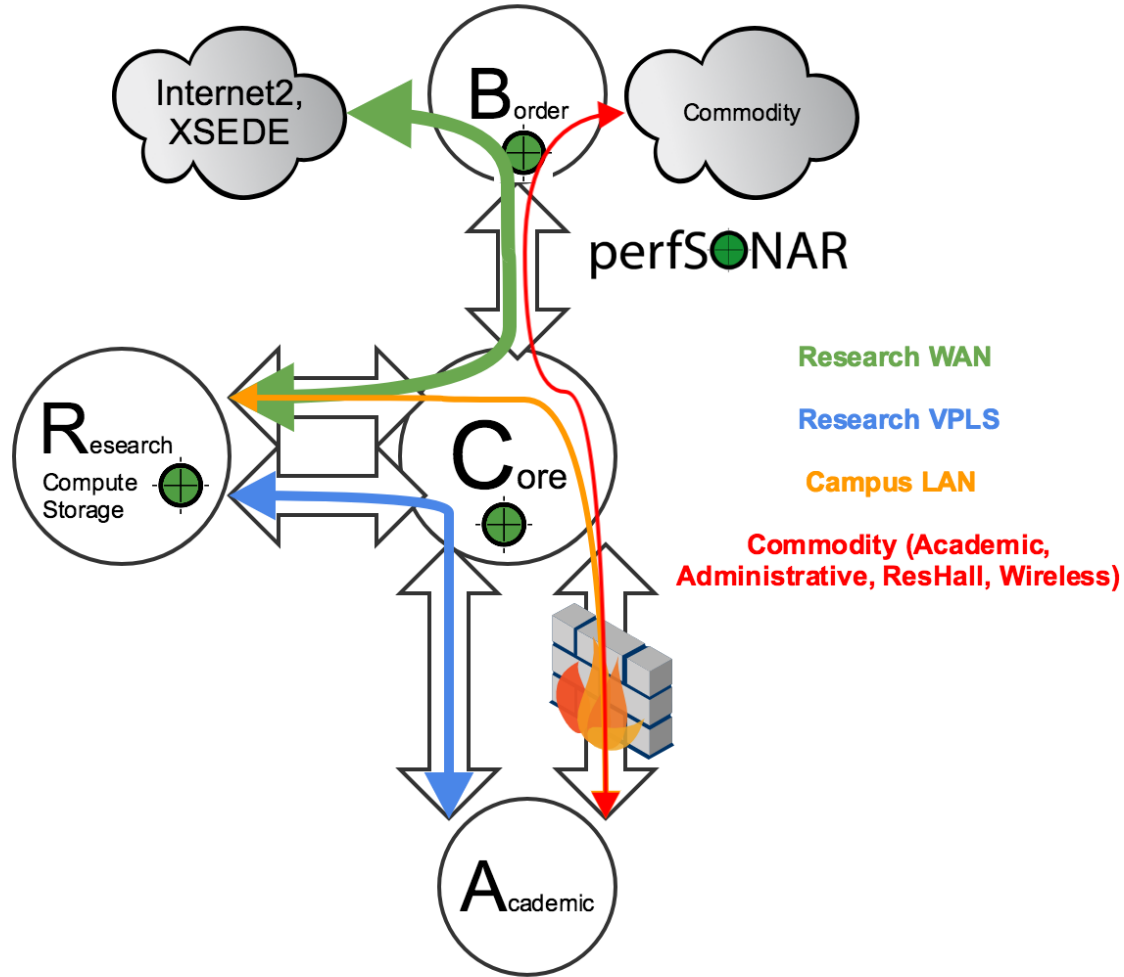
- Requires ACI to be in place – fast/low-latency networks, Science DMZs, SDN, perfSONAR, Advanced Computing, Storage, Data Transfer Nodes (DTNs), and a 100 Gb/s connection to the outside world
- Resources will be distributed across New Brunswick/Piscataway, Newark, and Camden
- Tiered Storage solution that will include HIPAA/FISMA compliance
- Designed to be a testbed to serve multiple research needs (including industry)
- Elastic in the sense that we will grow and shrink into cloud resources based on demand and job type (OSG, Google, Amazon, Azure)
- Couples NSF funded national projects and commercial cloud services directly into our environment creating a one-stop shop for the researcher
- Designed to be a Plug-N-Play environment with deployment of DTNs (FIONAS)
- As network is upgraded, more research groups will be added

Penn State/Pennsylvania Applications

- Meteorology
- Climate Modeling
- Bioinformatics
- Biochemistry and Molecular Biology
- Cryo-electron microscopy
- Astronomy

Penn State Science DMZ

Penn State Science DMZ and Research Network



Collaboration

- Formed networking group as part of the collaboration
 - Mailing list and biweekly meetings
 - Network engineers and project participants from Rutgers, Penn State, KINBER and NJEdge
- Activities
 - Technical details for implementing Data DMZ
 - General exchange of technical information, best practices

Project Status

- It is hard !
 - New network connections, for example, to the specific VDC Hub location at each campus
 - Accommodate infrastructure upgrades
 - New machine room at Penn State
 - NJEdge backbone upgrade
 - PennREN move to new colocation facility
 - Policy – who pays for the costs of the new connection if it wasn't included in the original grant
 - Identifying initial testing platforms

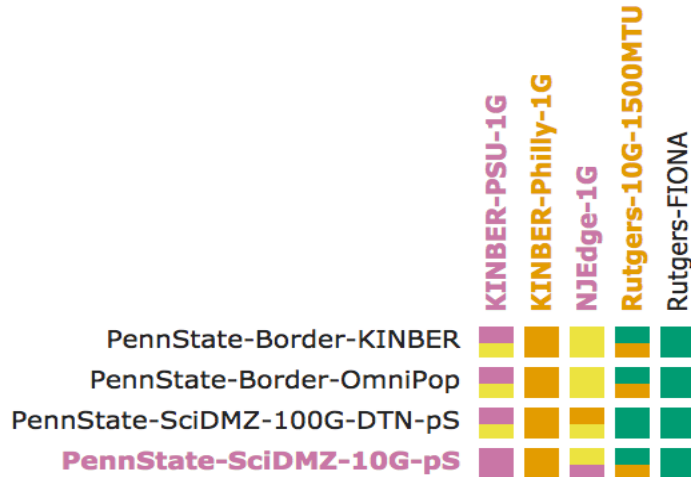
Observations

- Baseline network tests
 - Existing perfSONAR platforms, eventually include FIONA platforms at each site
- Network connectivity prior to connecting DMZ
 - Not optimal
 - 3 Internet2 providers (OmniPoP, Drexel and MAGPI)
 - Initial VDC Hub traceroute –
 - Return path used the commodity network connection and not the regional network or Internet2 connection

Initial Data DMZ perfSONAR Dashboard

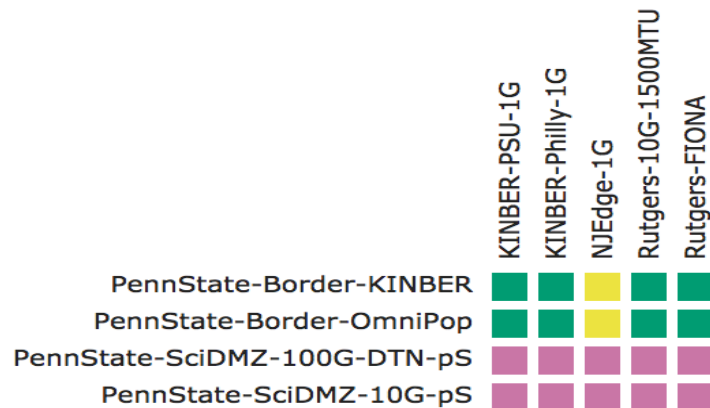
■ Throughput \geq 2000Mbps
 ■ Throughput $<$ 2000Mbps
 ■ Throughput \leq 500Mbps

! Found a total of 6 problems involving 5 hosts in the grid



■ Number of Paths is \leq 1
 ■ Number of Paths is \geq 1
 ■ Number of Paths is \geq 2

✓ No problems found in grid



What's Next

- Leverage the data DMZ concept to develop a “regional research platform”.
- Find science drivers among the initial hub and spoke sites associated with the VDC.
- Understand opportunities for expanding to other sites.
 - In New Jersey and Pennsylvania using the regional networks for connectivity.
 - In the region
- Connect to the national fabric.

Q&A - Discussion

Contact Information:

Wendy Huntoon (huntoon@kinber.org)

Barr Von Oeshen (barr.vonoehsen@rutgers.edu)

Ken Miller (kdm193@psu.edu)

Jim Stankiewicz) (jim.stankiewicz@njedge.net)

