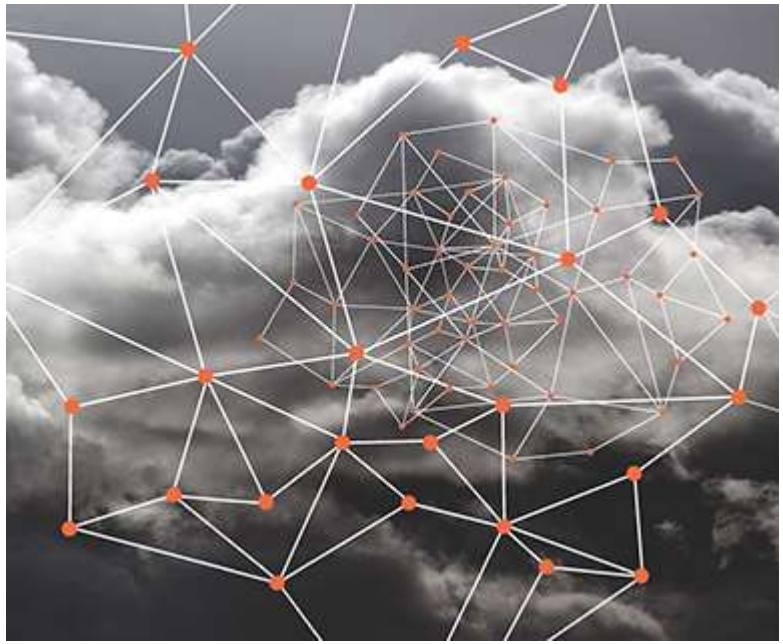


Elephant Flows in the Room: ScienceDMZ Nationally Distributed

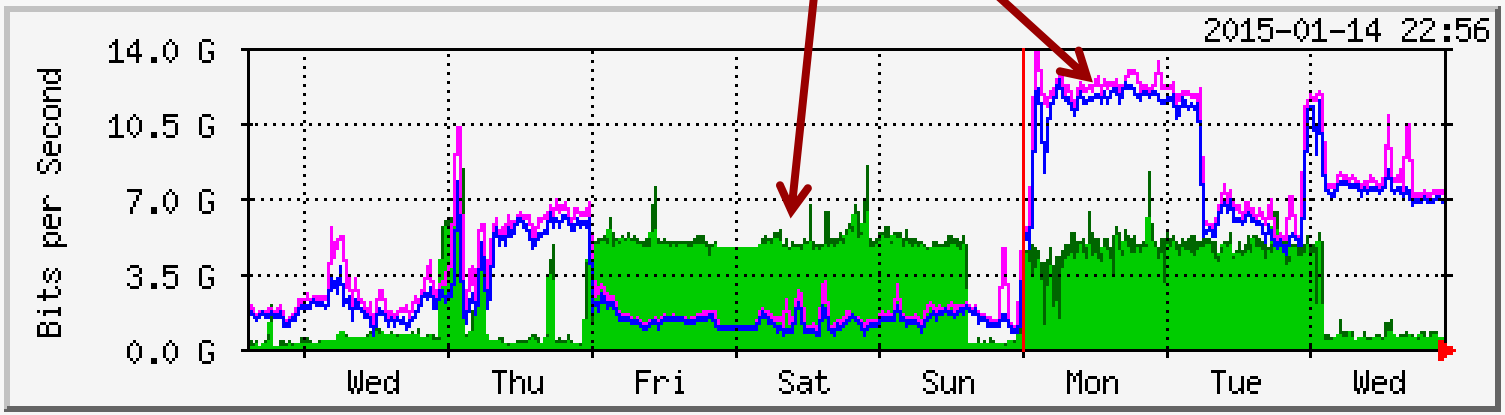
Do you know what your campus network is *actually* capable of ?

(i.e. have you addressed your elephant…)

Why ?

The network has been purpose built for research !!

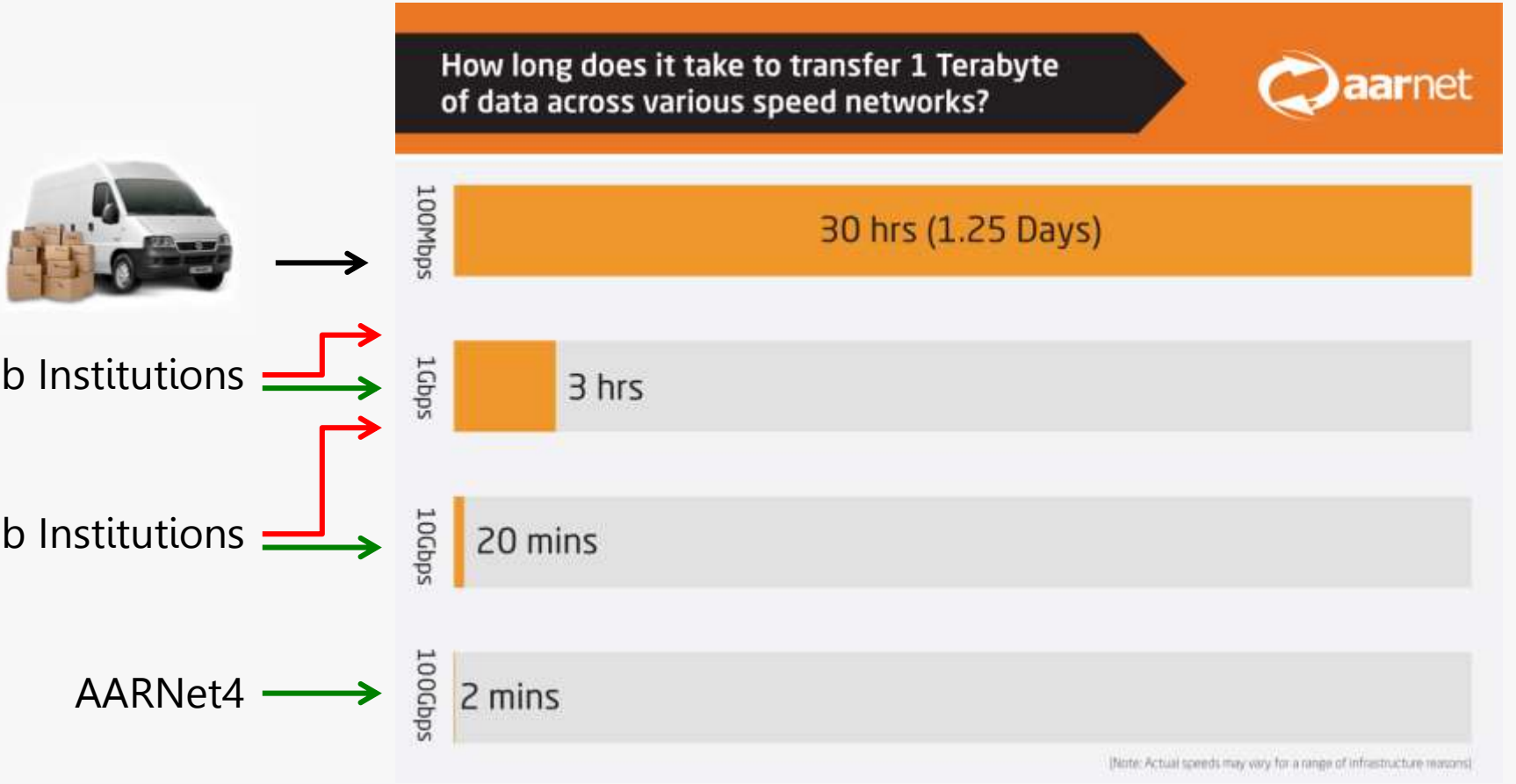
“Elephant” Flows



Our Aim

To increase the expectations of researchers from every institution to move large volumes of data.

Transfer Time: One Terabyte



One Terabyte on 10Gb *actual* transfer time = ~~**20 min**~~ **2-3 hrs**

Transfer Time: One Petabyte

1Gb Institutions

10Gb Institutions

AARNet4

Transfer Time

	1 TB	10TB	100TB	1PB
100 Mbps	1 day	10 days	3 months	3.4 years
1 Gbps	2.5 hours	1 day	10 days	3 months
10 Gbps	13 min	2.5 hours	1 day	10 days
100 Gbps	< 2 min	13 min	2.5 hours	1 day

Institutions

AARNet4

What's the problem?

- Security policies designed for enterprise, not science
- Poor choice of transfer tools
- Poor visibility of network, both national and international
- Ingrained workflows / behaviours
- Shortage of skills
- Poor visibility on user practice (i.e. "who knows what they're doing!")
- No support

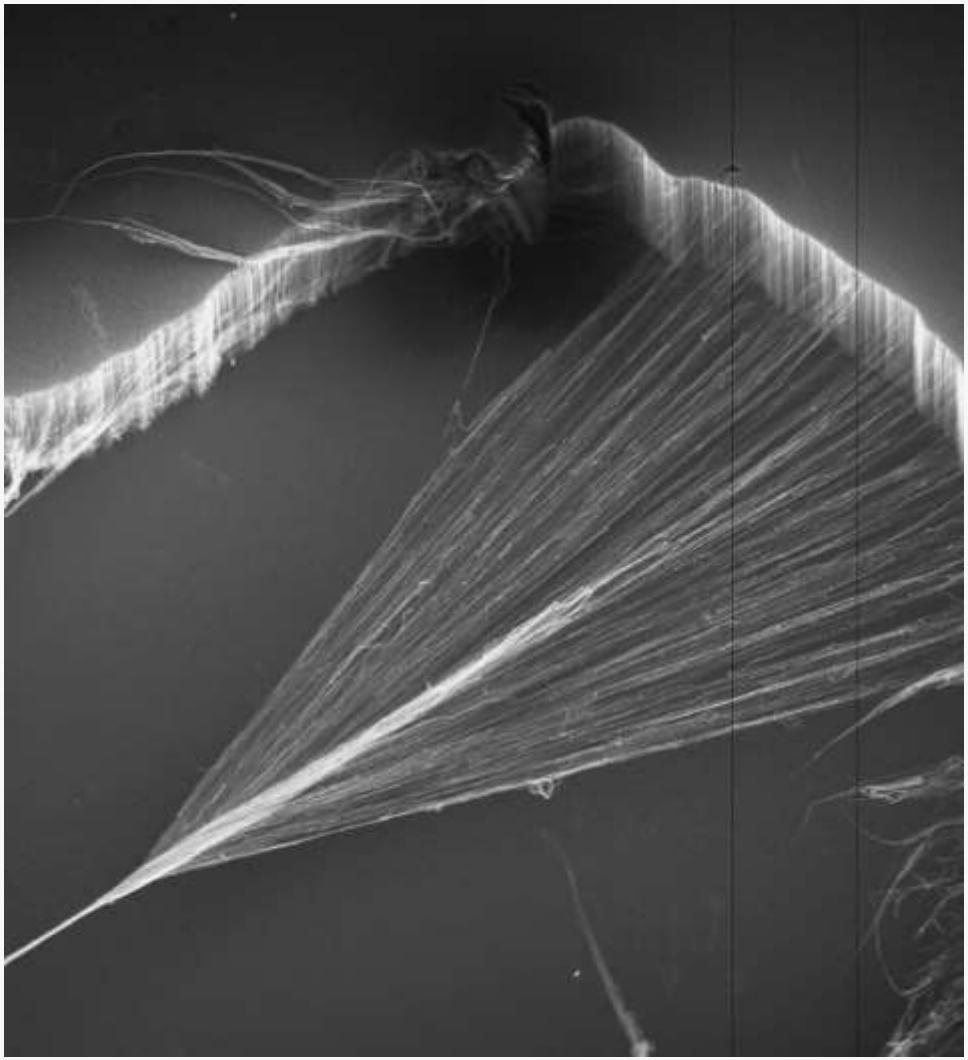
A big clue: Use of Sneakernet

Moral: There's more to large data transfer than the size of your pipe.

SO, what exactly is this “Science DMZ” and how will it help?

Overview

The Science DMZ is a portion of the network, built at or near the campus or laboratory's local network perimeter that is designed such that the equipment, configuration, and security policies are optimised for high-performance scientific applications rather than for general-purpose business systems or “enterprise” computing.



Carbon nanotubes being spun to form a yarn, CSIRO, source: https://en.wikipedia.org/wiki/Carbon

Overview

A Science DMZ integrates **four key concepts** into a unified whole that together serve as a foundation for this model

Architecture	• Manage “science” network independently
Security	• Tailored for few large science flows
Monitoring	• Characterise and set network performance
Data Transfer	• Dedicated tuned systems and tools

Overview

Architecture

- Manage “science” network independently

- Separates “Data Intensive” resources from the main campus network
- Allows more appropriate security methods to be applied for large flows
- Accommodates a range of scalable connectivity options, including redundancy, VPNs and additional wavelengths

Overview

Security

- Tailored for few large science flows

- Typical firewalls support many small flows rather than few large ones – the small buffers lead to significant packet loss, and subsequently very poor performance
- As most large transfers are between few resources, use Router ACLs instead of very expensive “large buffer” firewalls
- Separating elephant flow security from campus security is a relatively inexpensive win-win for both science and campus

Overview

Monitoring

- Characterise and set network performance

perfSONAR Nodes:

- Allows proactive monitoring of critical science pathways
- Set expectations for end to end network performance
- Uses a dedicated mesh of tools spread across the national research network, with a dashboard for overall visibility
- Creates a ready deployed fault-finding mesh to assist rapid targeting for network errors

Overview

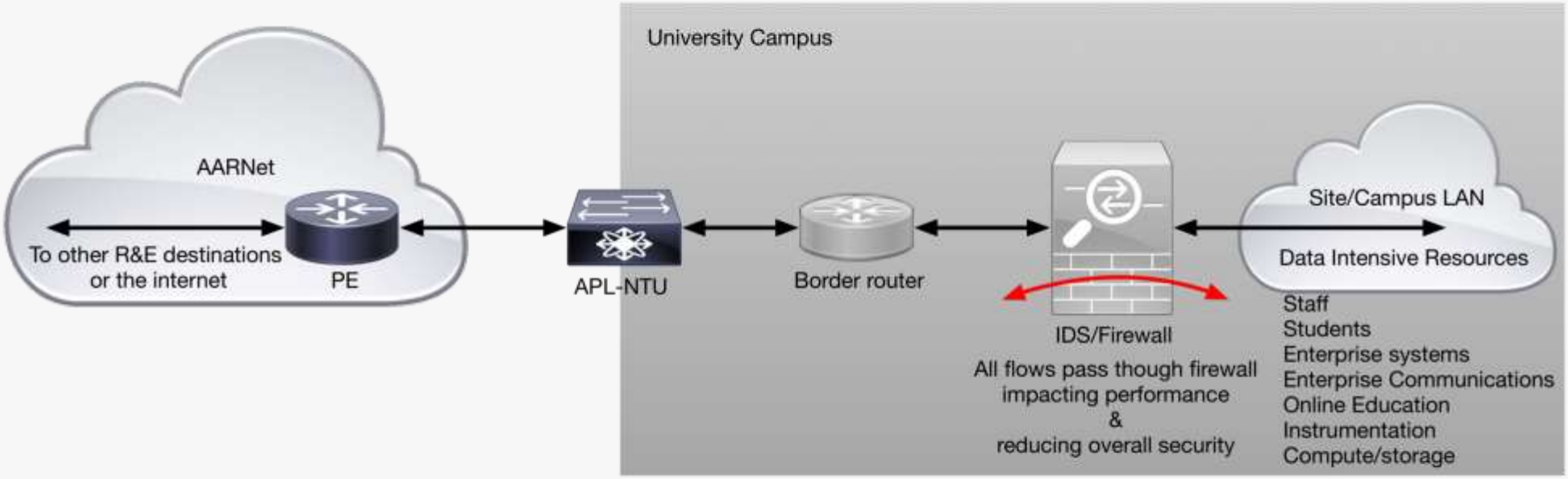
Data Transfer

- Dedicated tuned systems and tools

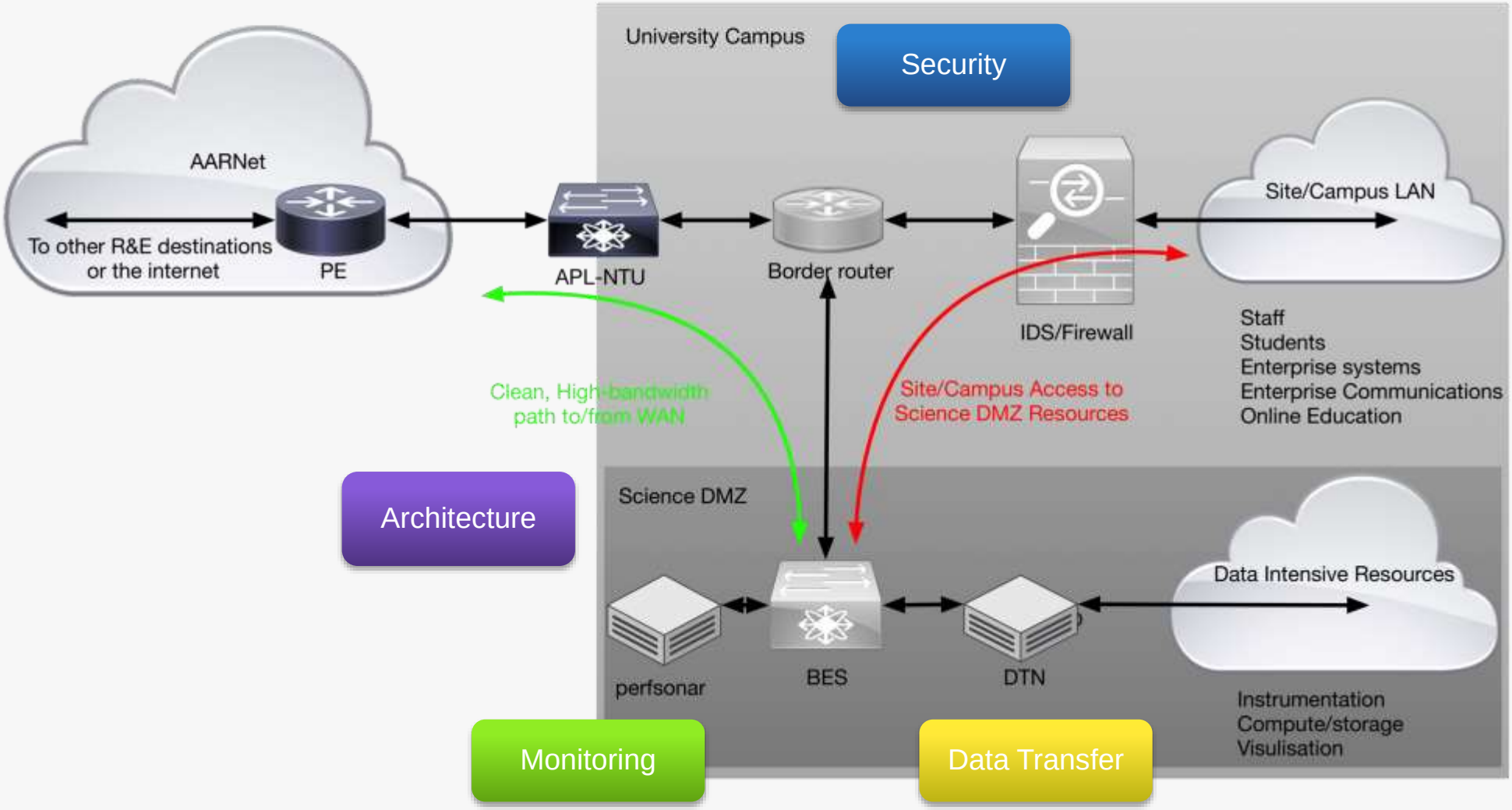
"Data Transfer Nodes" or simply "DTNs":

- Have access to local storage via either direct high-speed disk, a SAN or mounted high performance parallel filesystem (e.g. Lustre or GPFS)
- Have tuned high speed network interfaces, matched to the wan bandwidth.
- Have a collection of proven transfer tools which can sustain high data transfer rates over long latencies, e.g Aspera, Globus
- Do no general purpose computing tasks to mitigate security risks

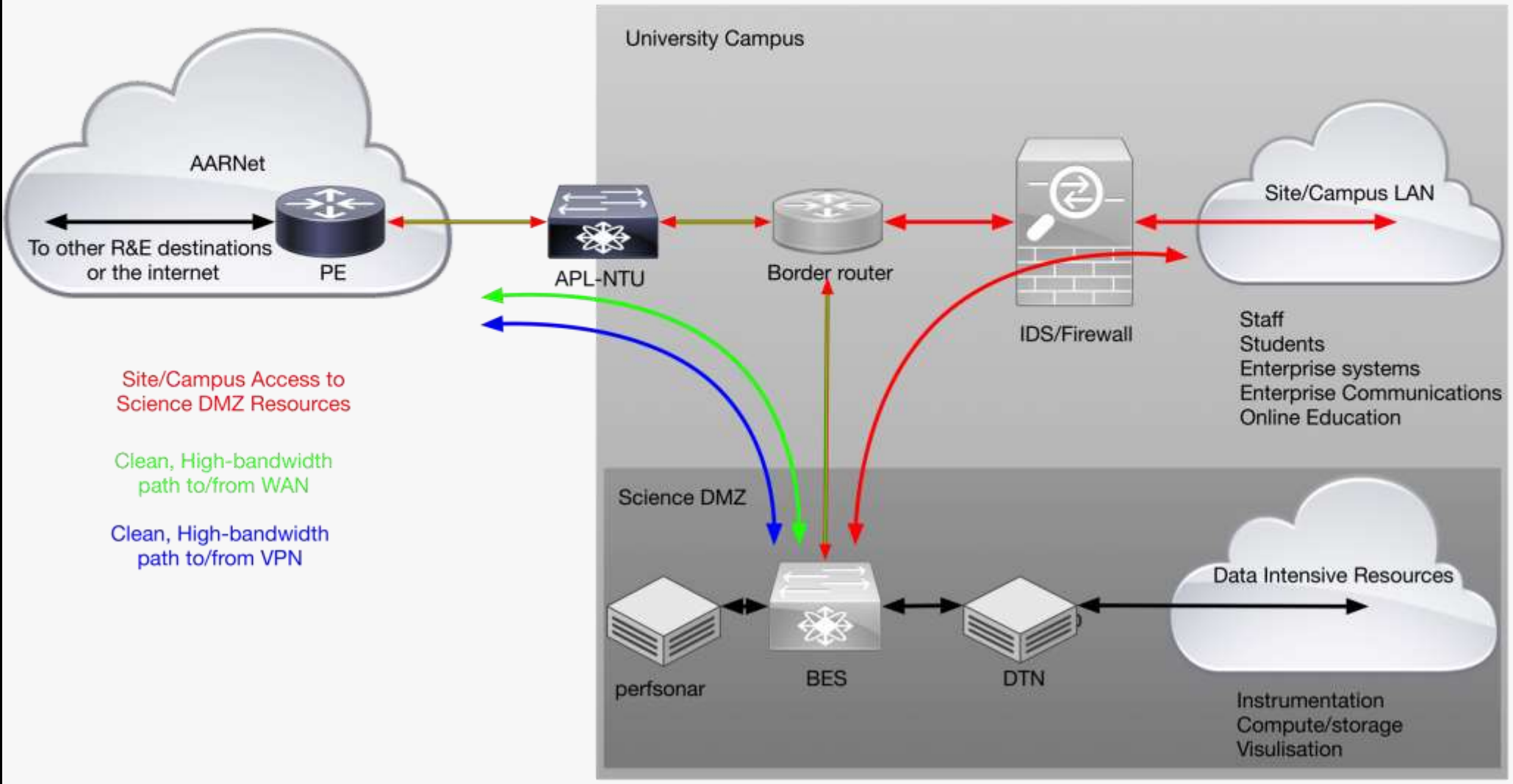
Example: Current Connection



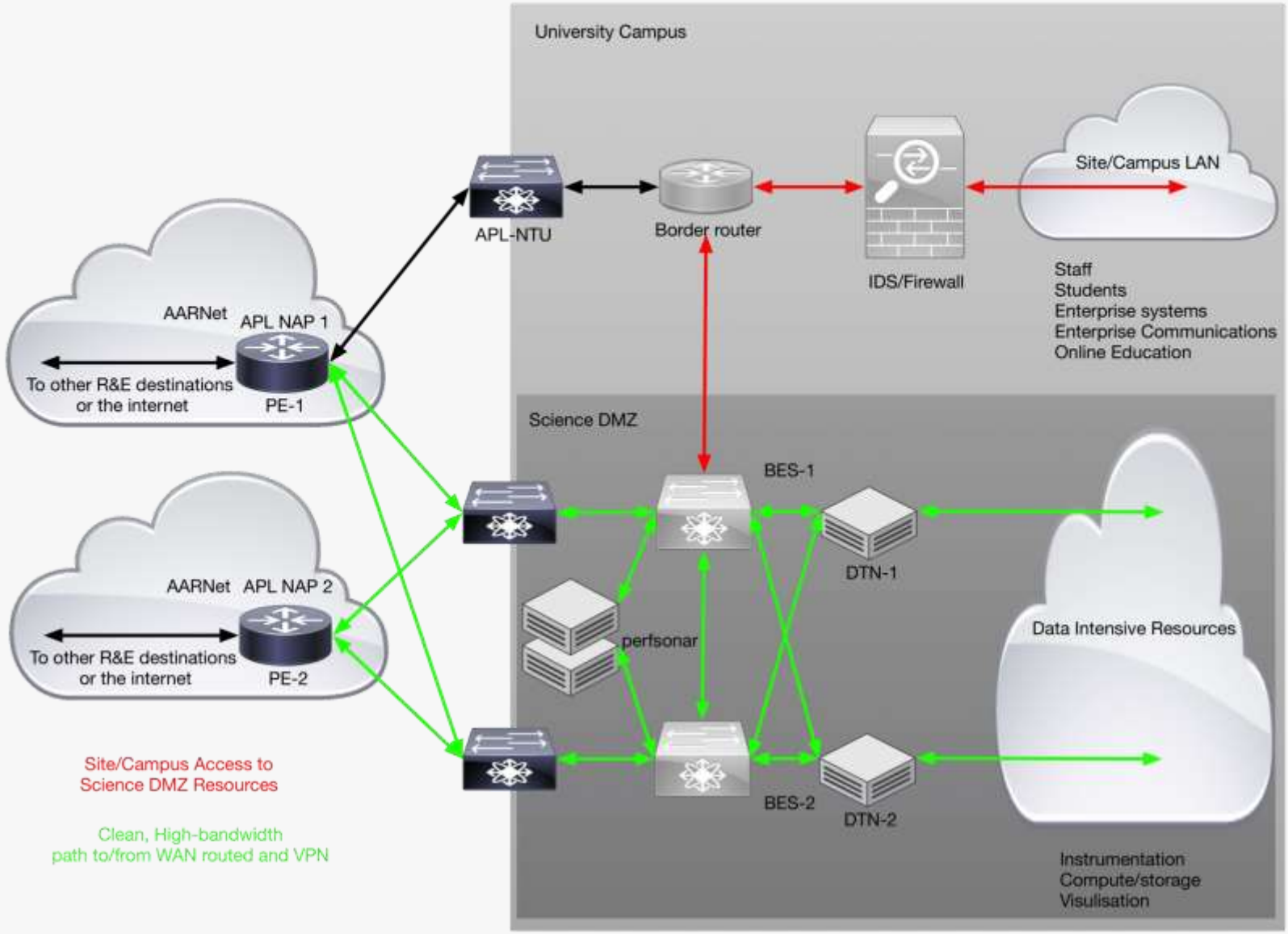
Example: Dependent Simple Connection



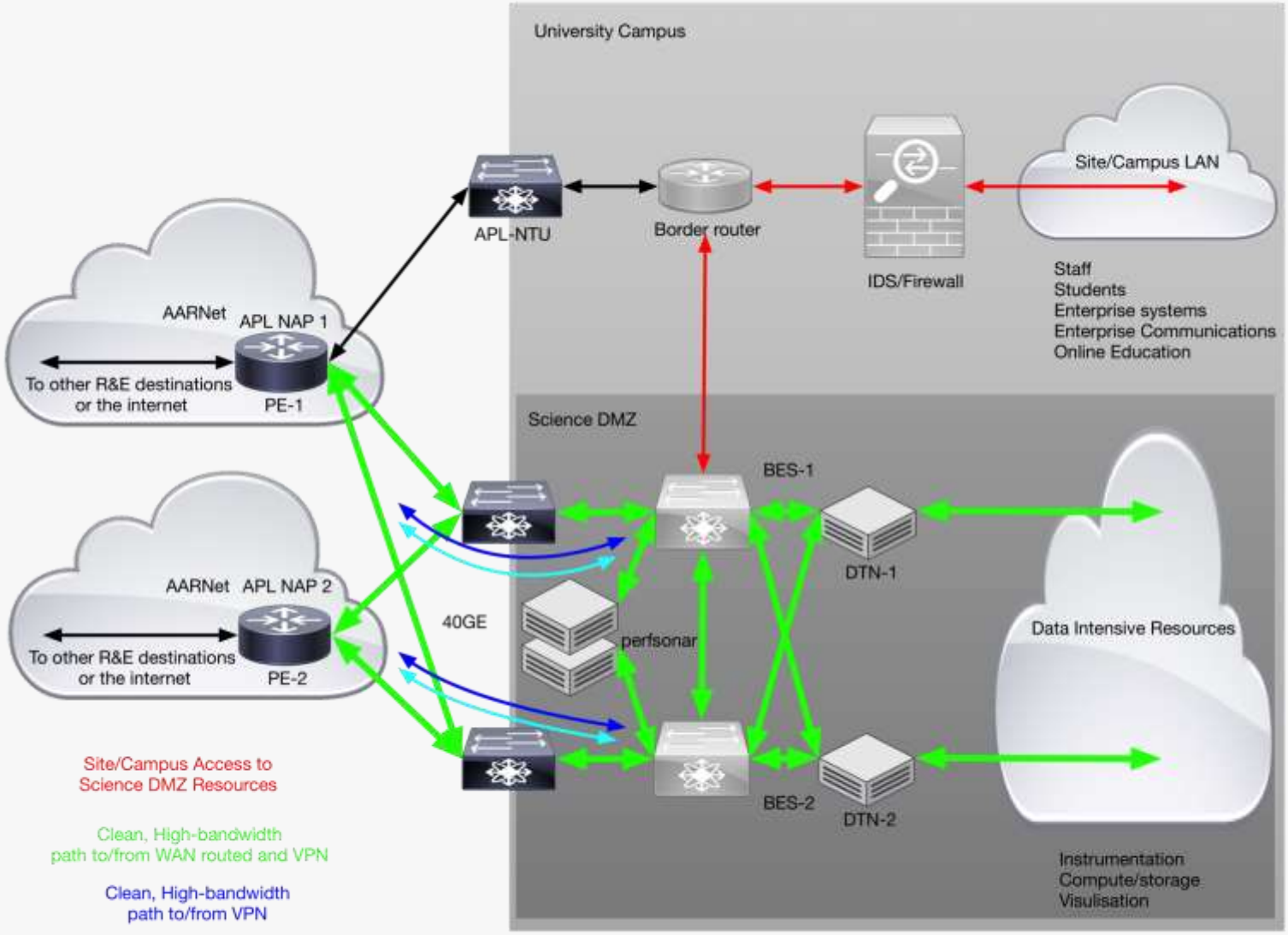
Simple SDN Connection



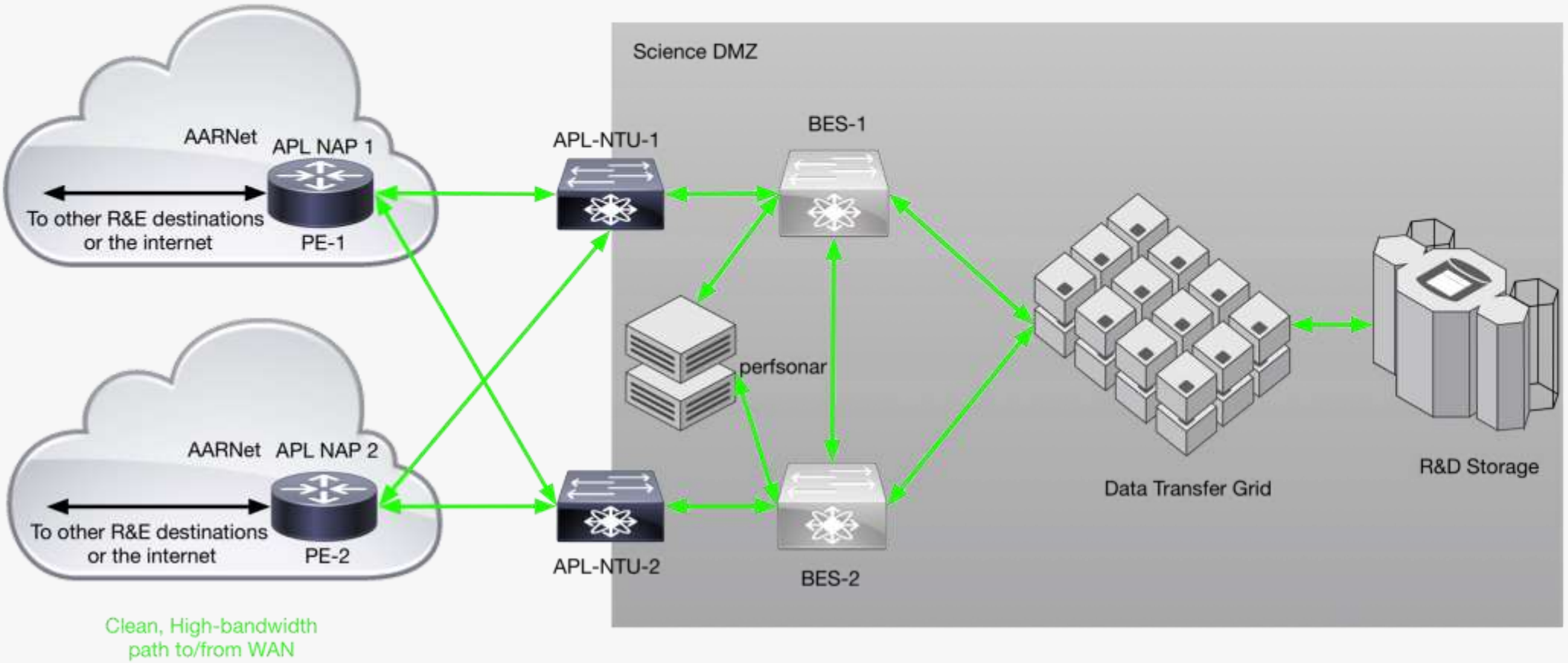
Redundant Standard Connection



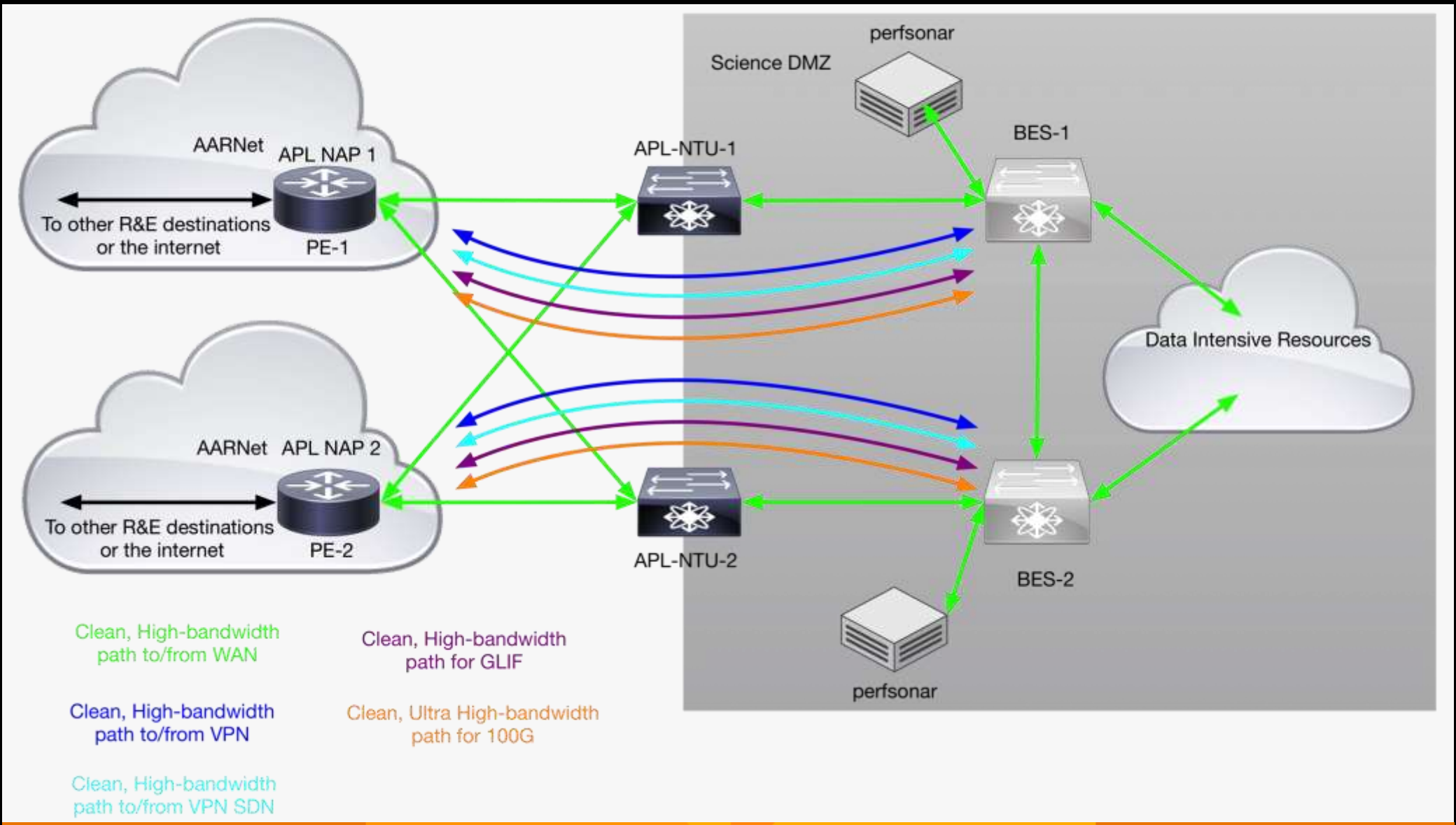
Redundant Standard 40G Connection



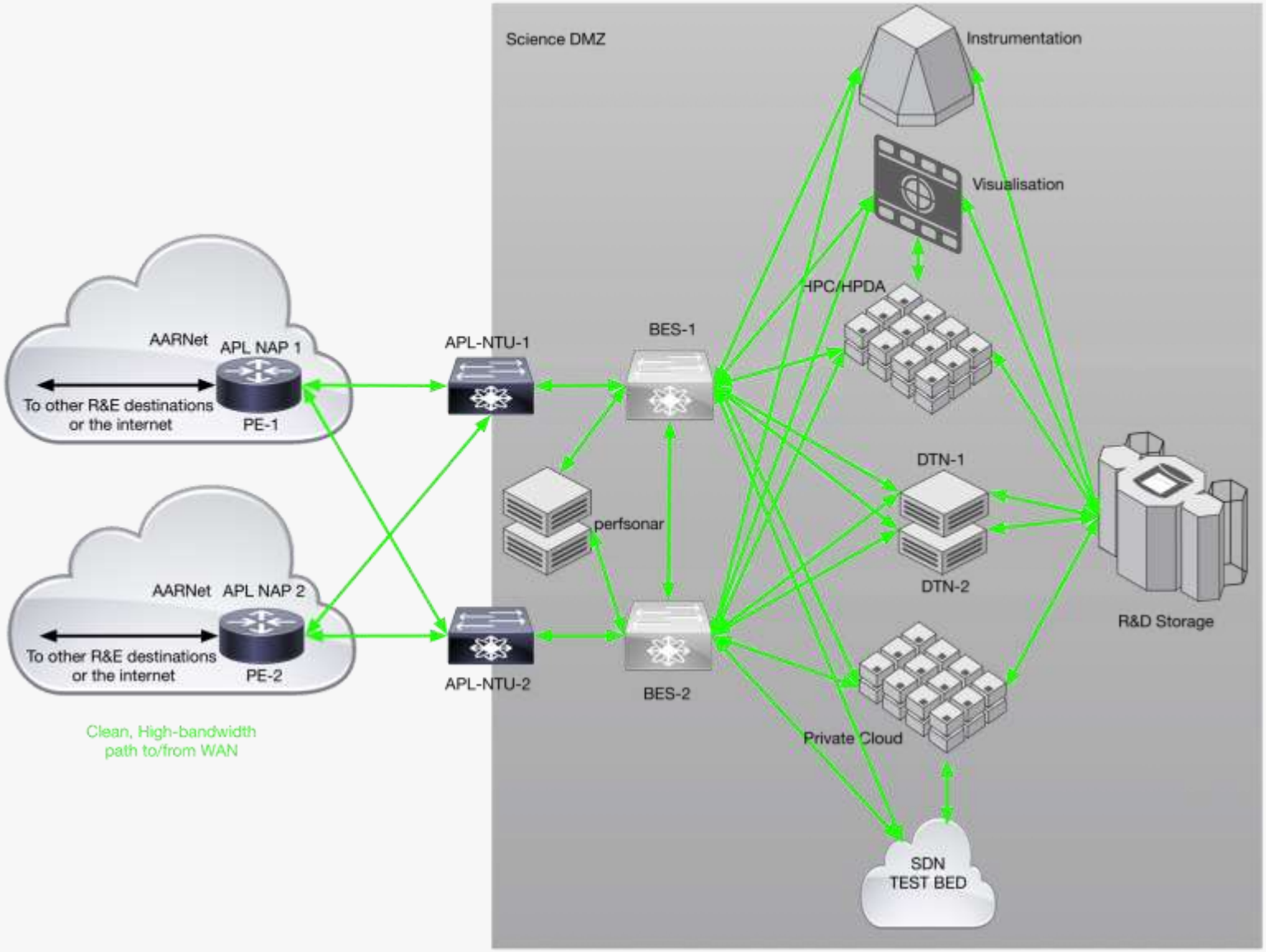
Redundant Big Data Connection



Redundant Complex Connection



On Campus Services Connection



aarnet

“Has Science DMZ made a difference?”

Demonstrated result

Internationally

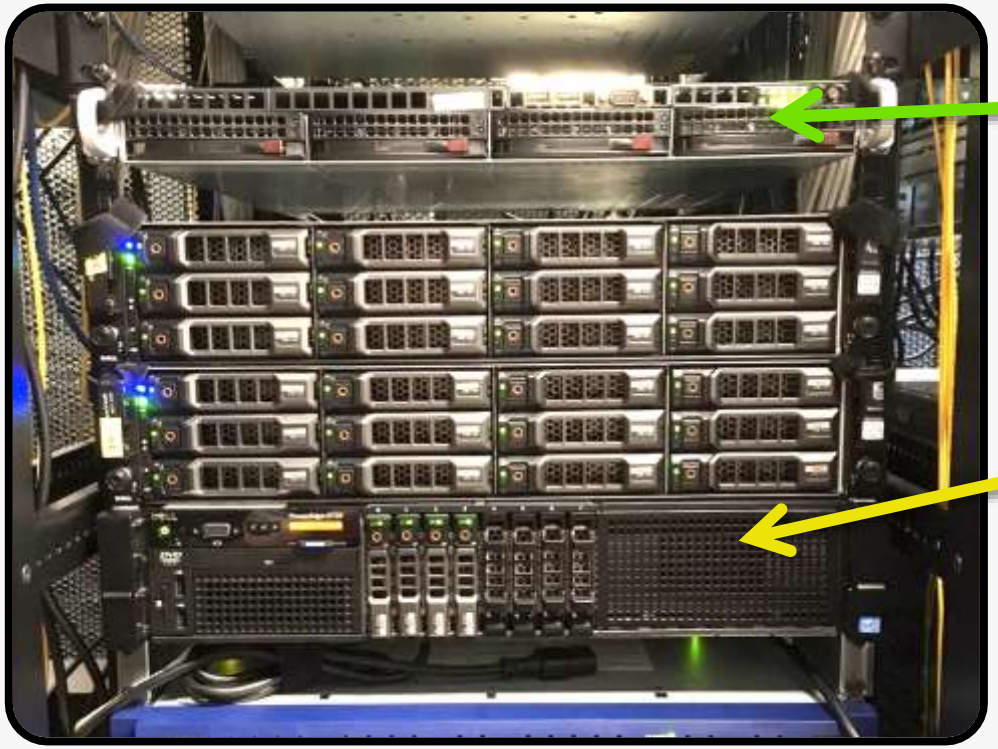
- ESNet approach to answering the Big Data Flow problem has refined the architecture, see fasterdata.es.net

In Australia

- Demonstrated significant performance improvement to line rate utilisation through robust deployment of DaShNet (RDSI) and national infrastructure
- Bottlenecks now *not* at Science DMZ capable end points, but often "the other end" or further up the stack.

Sender and Receiver capabilities are both critical.

Benchmarking Deployments



Monitoring

Data Transfer Node

2 x AARNet4 ScienceDMZ benchmarking deployments, capable of performing data transfer tests at 10Gbps

On Campus Testing Gear

1G Monitoring (Liva)

Portable 10G Data Transfer Node

Portable testing gear you can deploy at various locations across the campus to test against the benchmarking gear.

aarnet

"Is Science DMZ good for *my* university?"

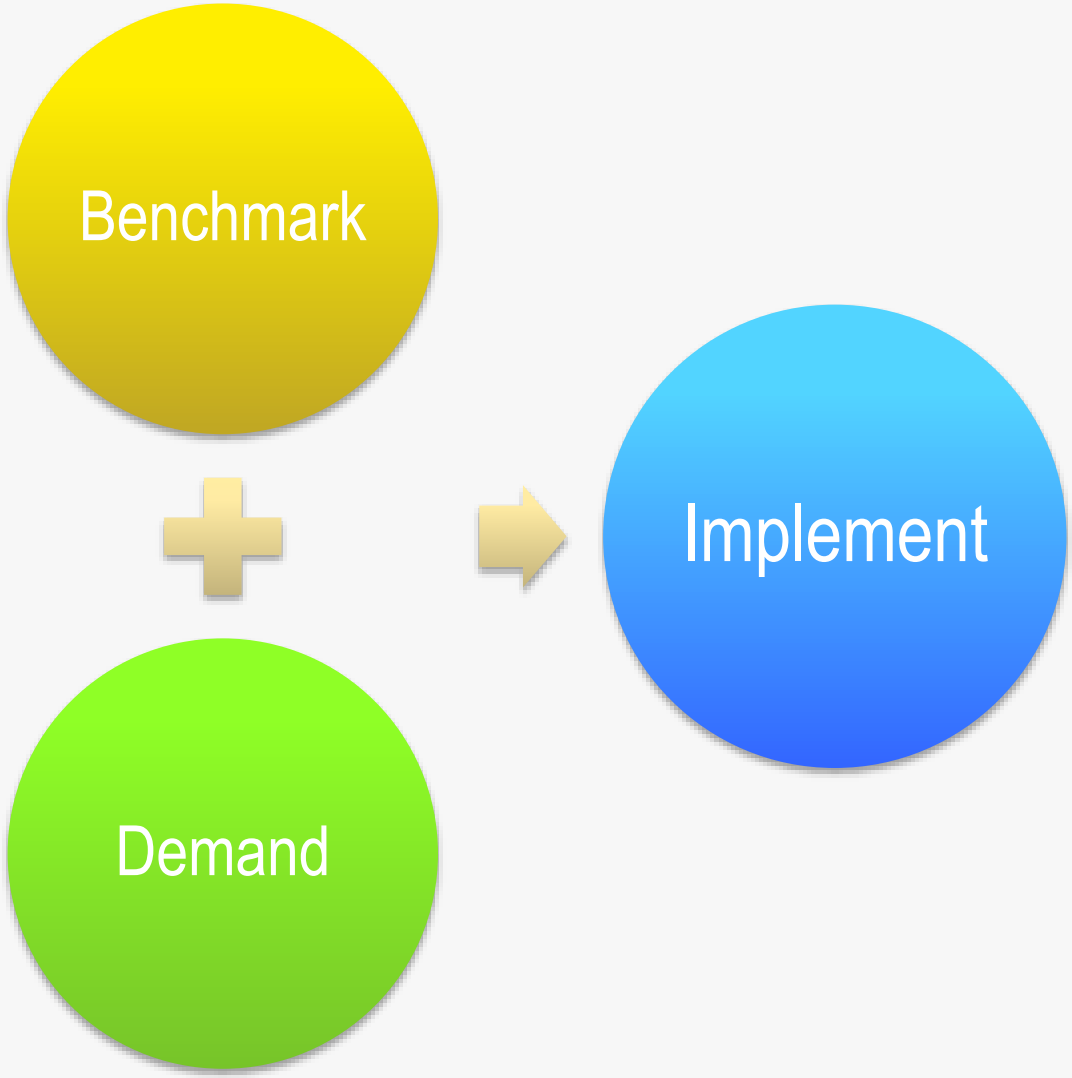
(Hint: the answer is going to have a "yes" in it somewhere …)

Do you qualify?

Answer "YES!" to any of these?

- ✓ Using or delivering national research infrastructure
- ✓ Data intensive research disciplines (established and emerging)
- ✓ General awareness of power users
- ✓ General lack of awareness of data handling techniques
- ✓ NetFlow data identifying "sleepers"
- ✓ Specific requests for more "network performance"
- ✓ No real testing done on border capability

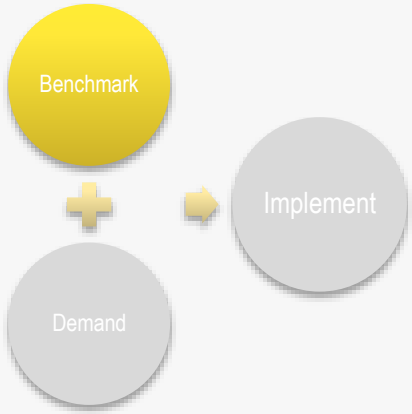
The Plan



Diligence on Benchmarking

In a coordinated manner with us:

- Test from border / faculty against Benchmarking capability
- Deploy test perfSONAR nodes on known transfers routes
- Conduct user machine / instrument throughput tests from deep within campus architecture

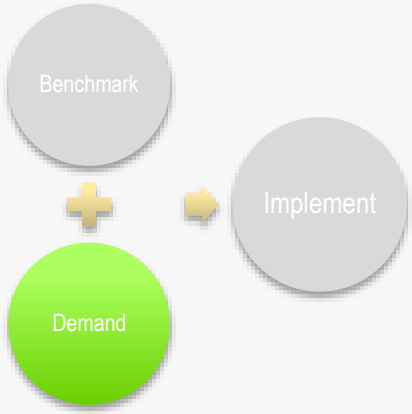


“How fast can you go?”

Diligence on Demand

In a coordinated manner with us:

- Capture transfer tools in use
- Assess NetFlow data
- Simply “ask around” research groups for indicative or expected performance

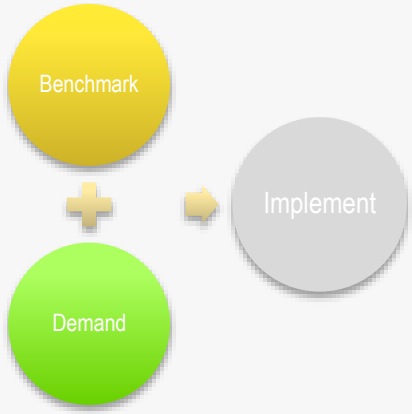


“Where’s your data?”

Outcomes from first two steps

Identifiable and validated outcomes:

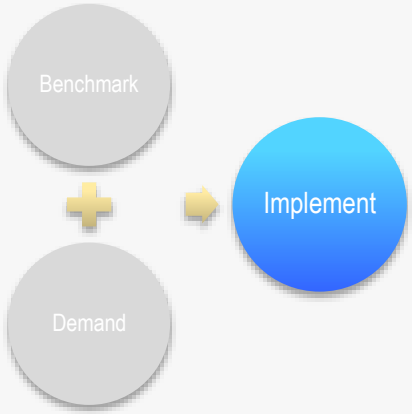
- Quantified performance across borders or known choke points
- Greater visibility on current and anticipated larger flows in/across/out of campus
- Greater visibility on discipline specific choice and capability of tools
- Better and proactive engagement with research.
- A demonstrated commitment to improve access to data



Follow Up Implementation

Identify Science DMZ architecture and/or components to address issues discovered then as needed:

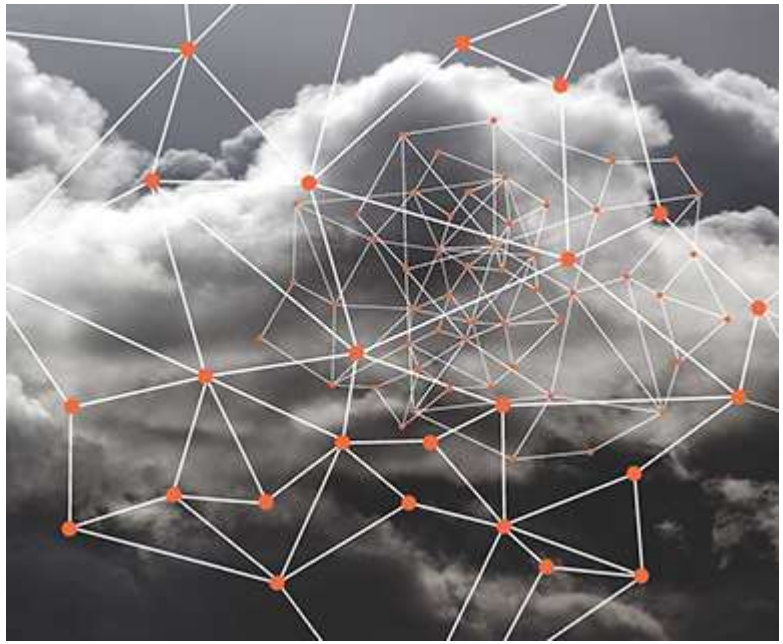
- YOU build and operate
- WE build, YOU operate
- WE build and operate = SDMZaaS
- Consult further with AARNet Enterprise Services for a broad multi-campus approach



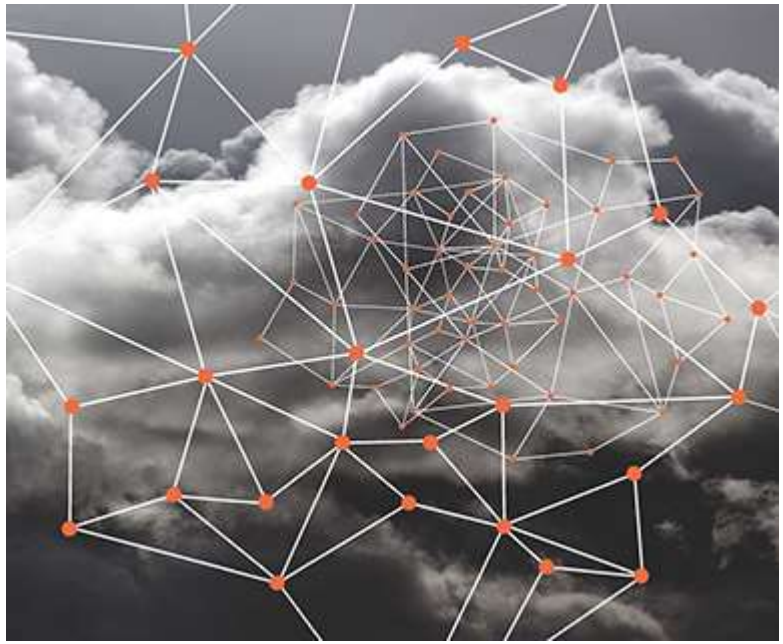
Result = Demonstrable follow through in support of research

Where to start?

Datamovers@aarnet.edu.au



THANKS ☺



SPARES

Summary

What's a Science DMZ?

Developed about six years ago by engineers at Energy Sciences Network (ESnet) and National Energy Research Scientific Computing Center (NERSC) the Science DMZ refers to an operationally-proven network architecture optimized for the transfer of large-scale scientific data.

The model includes recommended hardware devices, security policies, and network performance software which together provide the ideal environment for moving science data as efficiently as possible.

Summary

In practice, a Science DMZ creates an enclave on a campus network that is specially designed for science data (a vastly different data profile than a campus' enterprise applications).

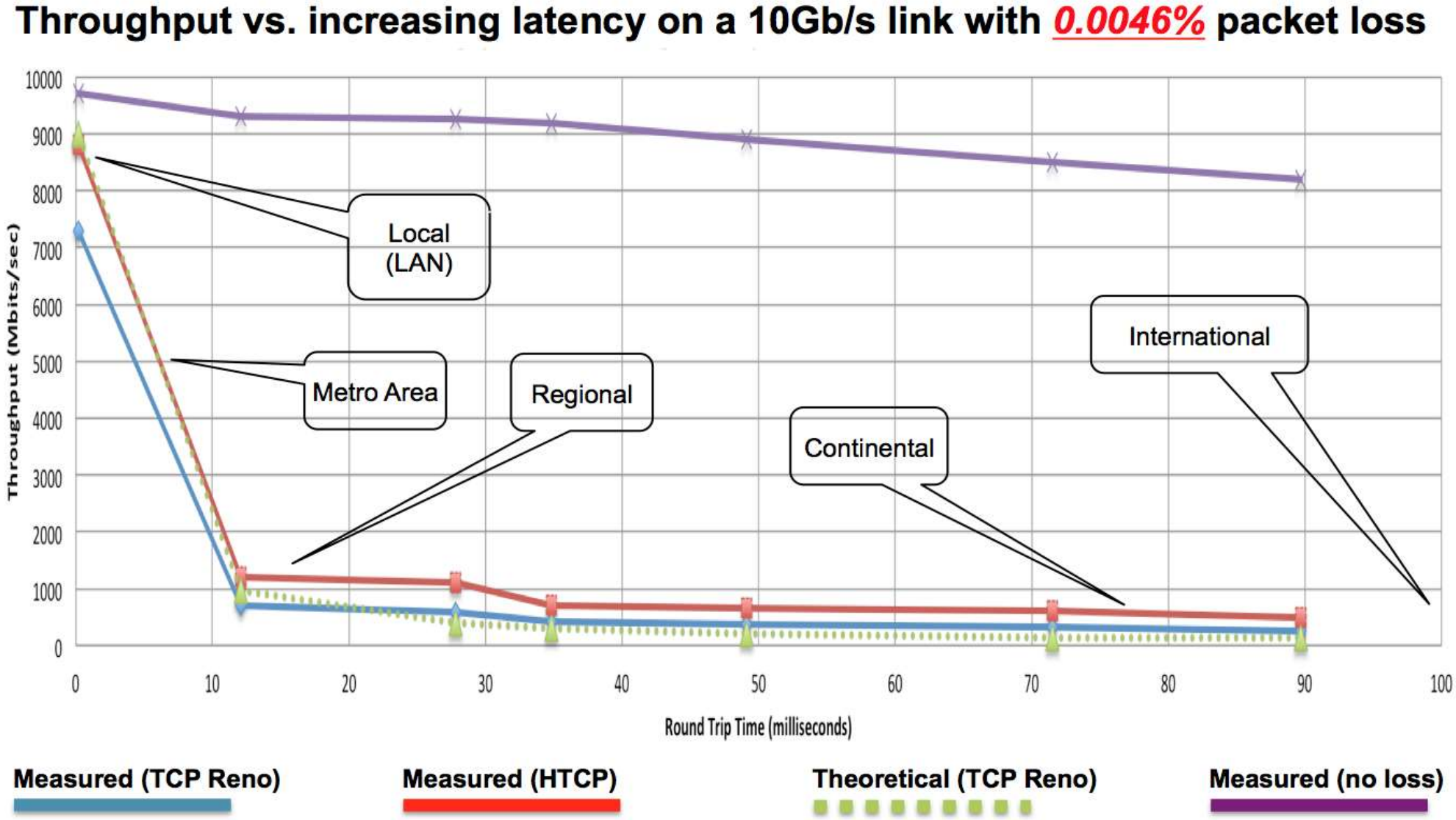
A Science DMZ recognizes that all the networked applications a university needs to run — whether for science or for business — have variable needs. By applying best practices for data management, the Science DMZ ensures the efficiency of science data and regular day-to-day university business applications is not impeded.

Overview

The Science DMZ Model addresses **several key issues** in data intensive science, including:

- Reducing or eliminating the packet loss that causes poor TCP performance.
- Implementing appropriate security architectures and controls so that high-performance applications are not hampered by unnecessary constraints.
- Providing an on-ramp for local science resources to access wide area science services including virtual circuits, software defined networking environments, and 100 Gigabit infrastructures.
- Incorporating network testing, network measurement, and performance analysis through the deployment of perfSONAR.

Overview


Throughput vs. increasing latency on a 10Gb/s link with 0.0046% packet loss
Throughput (Mbits/sec)
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
10
20
30
40
50
60
70
80
90
100
Round Trip Time (milliseconds)
Local (LAN)
Metro Area
Regional
Continental
International
Measured (TCP Reno)
Measured (HTCP)
Theoretical (TCP Reno)
Measured (no loss)

aarnet

Which architecture is right for you?