Managing Data Intensive Challenges with a Science DMZ

SwitchOn Workshop – São Paulo
October 15-16, 2015

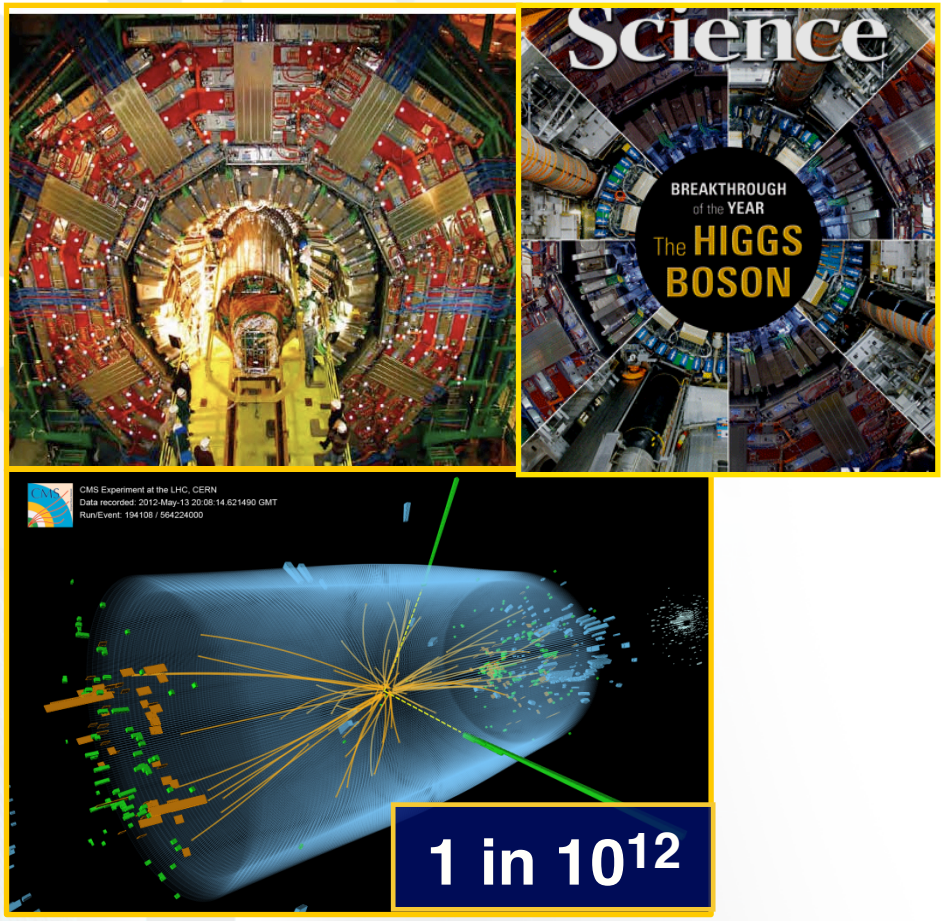
Julio Ibarra
Florida International University

Data Intensive Challenges

- Many Disciplines Need Dedicated High Bandwidth on Campus:
- Remote Analysis of Large Data Sets
  - Particle Physics

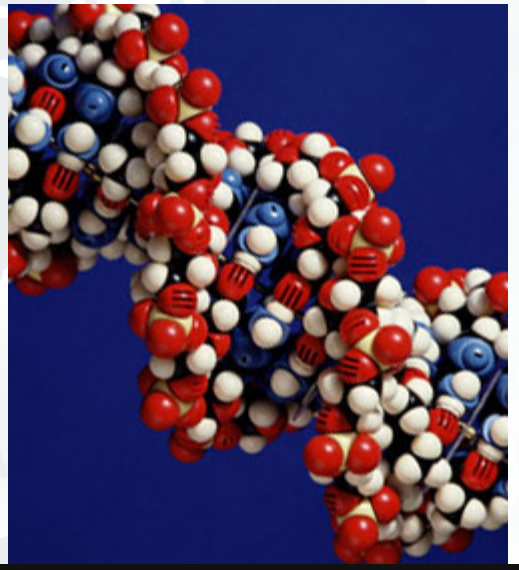
Particle Physics

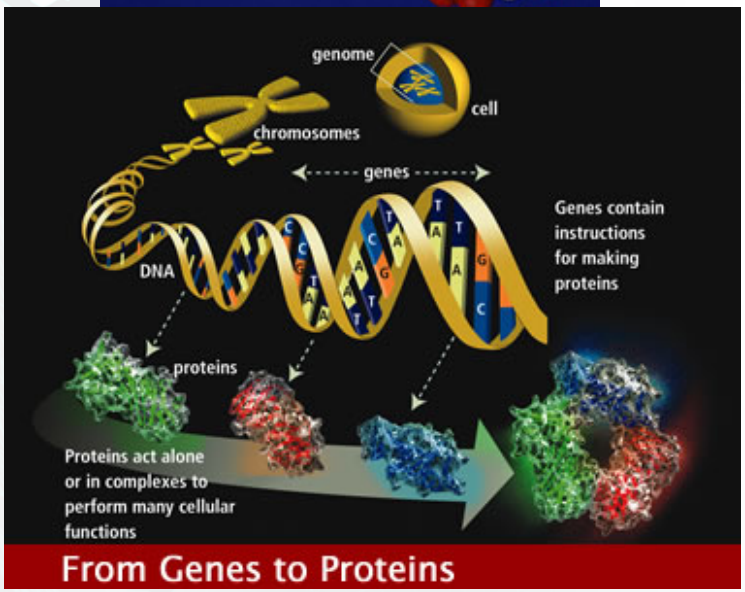
- Large Hadron Collider
- ATLAS / CMS
- ~10s of TBs of data transfers
- Data distribution across continents
- Distribution rates at 40+ Gbps
- Requirements increasing over time



Disruptive Genomics Technologies

- Data rates increasing exponentially
- Instrument data rate up 10x in 5 years
- Sequencing cost down 10x in 5 years
- New applications for genomics data as science improves





SWITCH ON
RESEARCH • COLLABORATION • EDUCATION

Campus Cyberinfrastructure

- Data intensive science requires
  - High-throughput networks
  - Available bandwidth
- Campus networks must support
  - Multiple business operations
    - Research and Education
    - Business enterprise applications
  - Security for protecting the data of the institution
- Data intensive science applications
  - Experience poor performance on campus networks
  - Campus networks must be engineered to support
    - The data movement requirements of data intensive science
- *How can campus networks be adapted to*
  - *Optimize flows from science applications, while*
  - *Not impacting the operation of the business enterprise?*

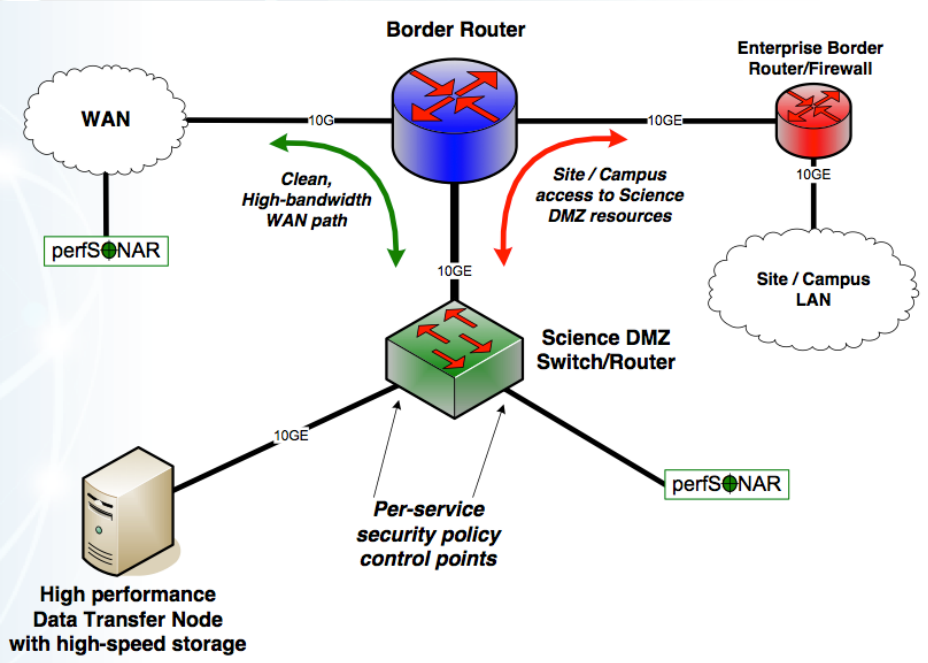
SWITCH ON
RESEARCH • COLLABORATION • EDUCATION

Science DMZ: Introduction

- Objective: Adapt the campus network to optimize science data intensive flows without impacting the operation of the production network
- The Science DMZ (S-DMZ) adapts a portion of the network that is engineered for science applications and does not not include support for general-purpose use
- The S-DMZ provides dedicated resources for wide-area data transfer

Science DMZ Components

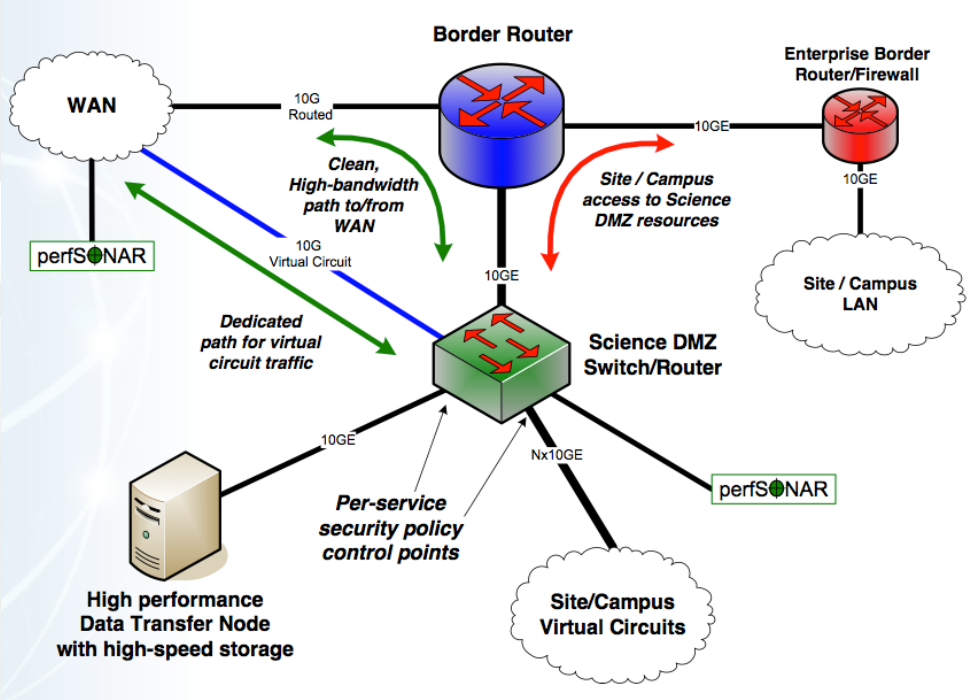
- Friction free connections for high-performance data transfer nodes to a Science DMZ switch/ router
- Border router partitions Science DMZ from campus LAN
- Supports per-service security policy control
- Multiple Measurement Points (S-DMZ, campus, WAN)



Source: Eli Dart, ESnet

Science DMZ with Virtual Circuits and Openflow

- Hybrid network services supported
  - OSCARS and other Inter-domain control planes
- Openflow/SDN adds support for virtualized control plane functions



Source: Eli Dart, ESnet

NSF Campus Cyberinfrastructure Program

- NSF Campus CI program has invested in the improvement and re-engineering at the campus level to leverage dynamic network services to support a range of scientific data transfers and movement
  - Campus Cyberinfrastructure - Network Infrastructure and Engineering Program (CC-NIE)
  - Campus Cyberinfrastructure - Infrastructure, Innovation and Engineering Program (CC*IIE)
  - Campus Cyberinfrastructure - Data, Networking, and Innovation Program (CC*DNI)

CC-NIE: Data Driven Networking Infrastructure for the Campus and Researcher Goals

- Network infrastructure improvements at the campus level
  - Supports a wide range of science data flows
  - Building a Science DMZ
- Enables national and global high-performance end-to-end access to dynamic network services
- Enables rapid, unimpeded movement of diverse and distributed scientific data sets and advanced distributed computing

CC-NIE: Network Integration and Applied Innovation Goals

- Supports the development and integration of innovative networking capabilities
- Enhances end-to-end network Cyberinfrastructure through integration of existing and new technologies and applied innovation
- Integration of networking protocols and technologies with application layer code and processes
- Transitions successful research prototypes in Software Defined Networking (SDN) to distributed scientific environments and campus infrastructure

Thank You!

Julio.Ibarra@fiu.edu

SWITCH ON
RESEARCH • COLLABORATION • EDUCATION