Ubiquitous Big Data:

Supporting the Proliferation of Big Data Experiments from the Data Center

Shigeki Misawa
BNL Scientific Data and Computing Center

BROOKHAVEN NATIONAL LABORATORY
a passion for discovery

U.S. DEPARTMENT OF ENERGY | Office of Science

Next Gen Big Data Experiments

- Expect dozens of unique instruments capable generating Big Data at BNL
- Proliferation of unique DAQ systems with limited data format/export capabilities
- Geographically dispersed on campus
- Each instrument may host multiple, independent experiments in a single year
- Individual sites may not be able to support necessary on line and off line compute infrastructure

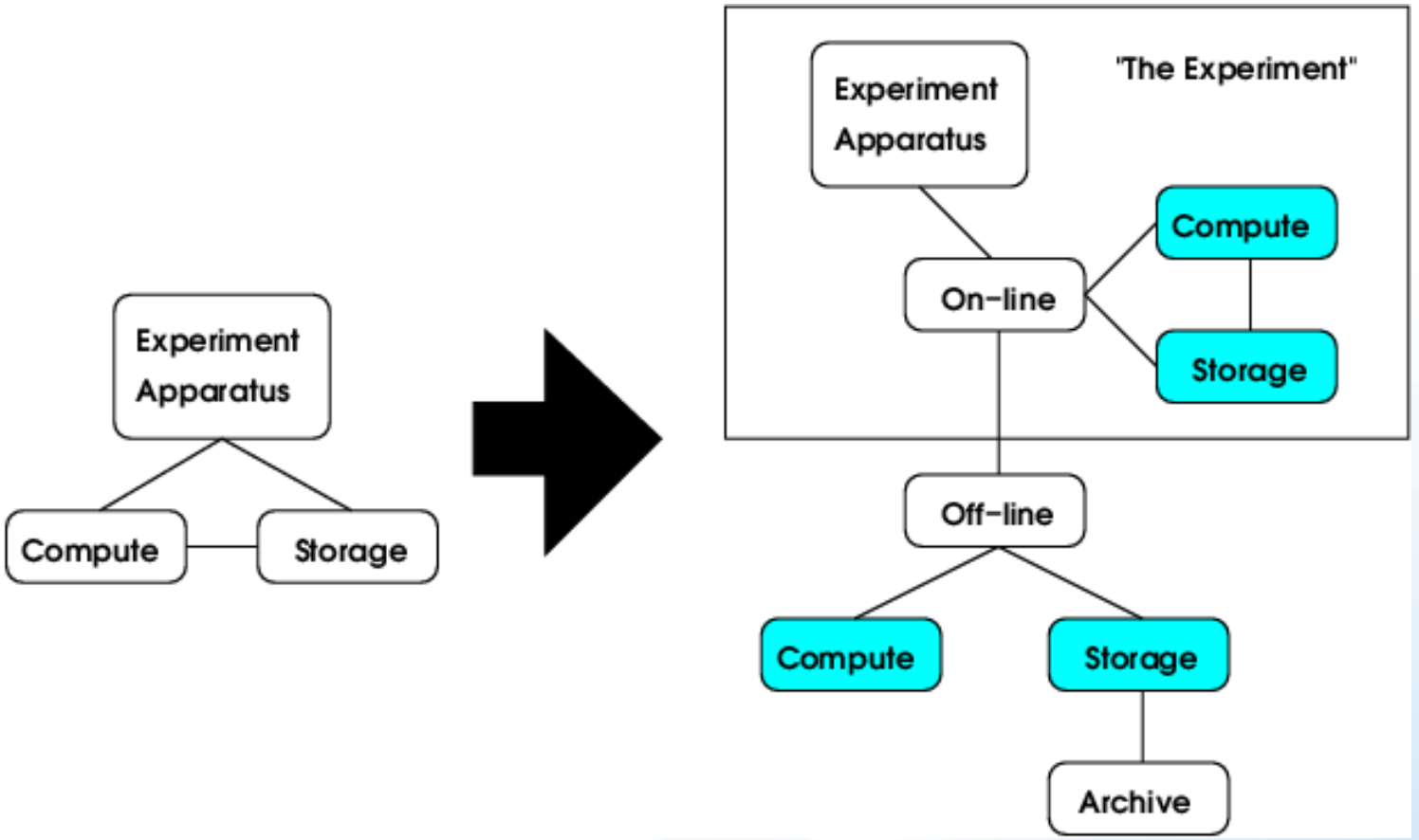
Next Gen Big Data Experiments

- Compute and storage requirements vary
  - By instrument
  - By the specific experiment run using the instrument
  - Over time
- By sharing resources, a group of next generation big data experiments may be able to collectively purchase, support, and operate the compute and storage systems they need.

Added Complications for Big Data (old and new)

- New data stewardship requirements (data life cycle management)
- Data confidentiality issues
- Fundamental changes in compute and storage technology are occurring at this time.
  - Grid/Cloud computing
  - Virtualization/Containerization
  - Storage technologies (software and hardware)
  - Heterogeneous computing (GPU, FPGA)
  - Software defined networks (SDN)

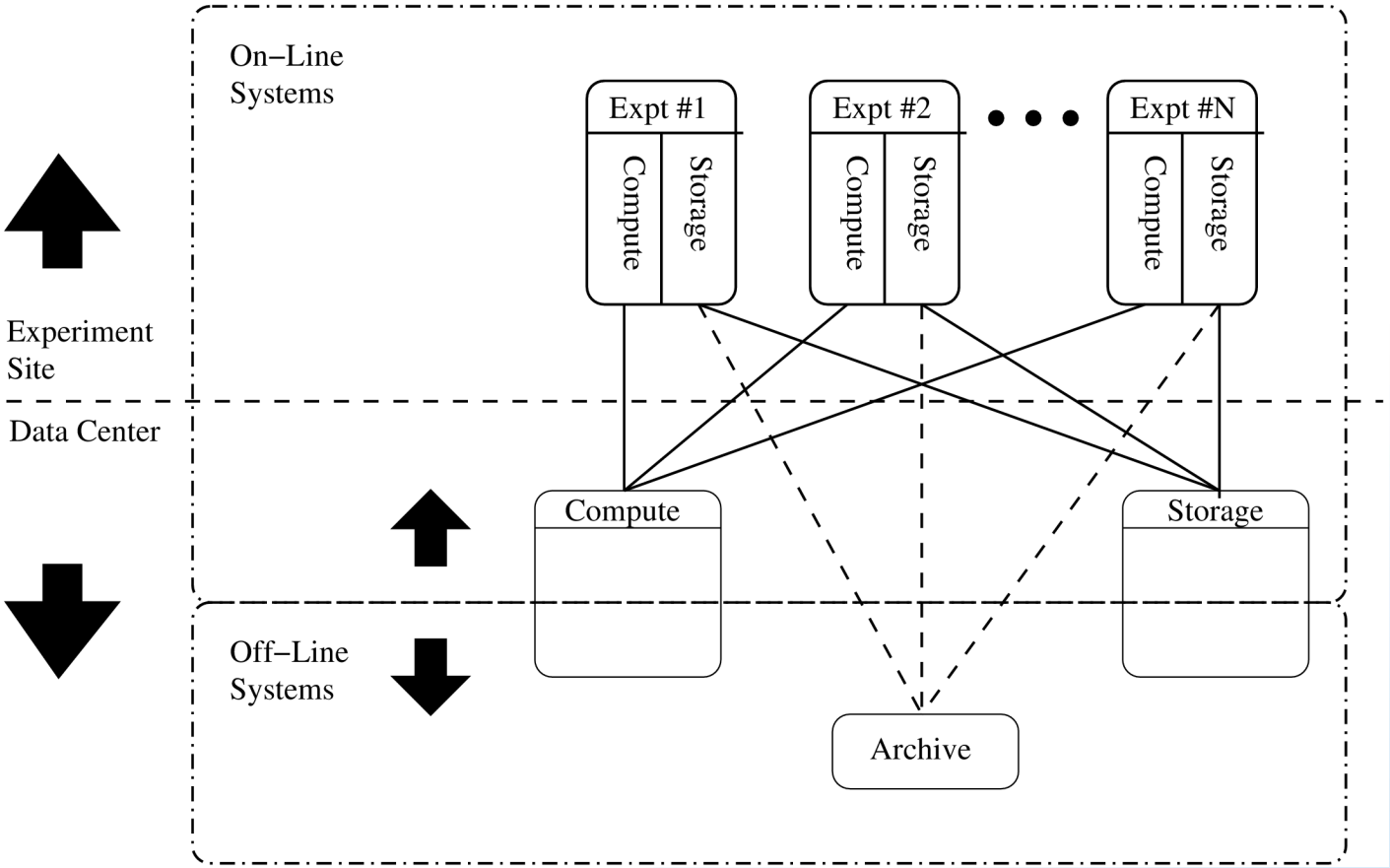
Revisiting Scaling of Experiments



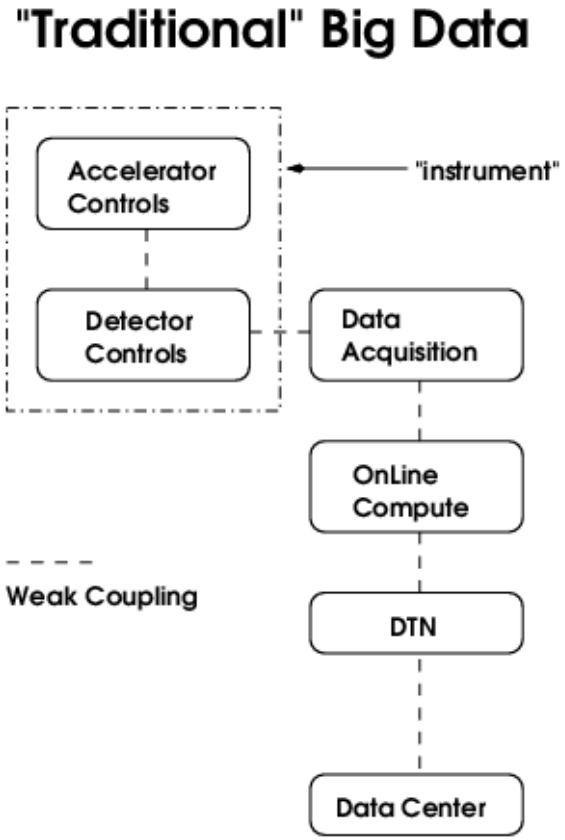
Supporting Next Gen Big Data

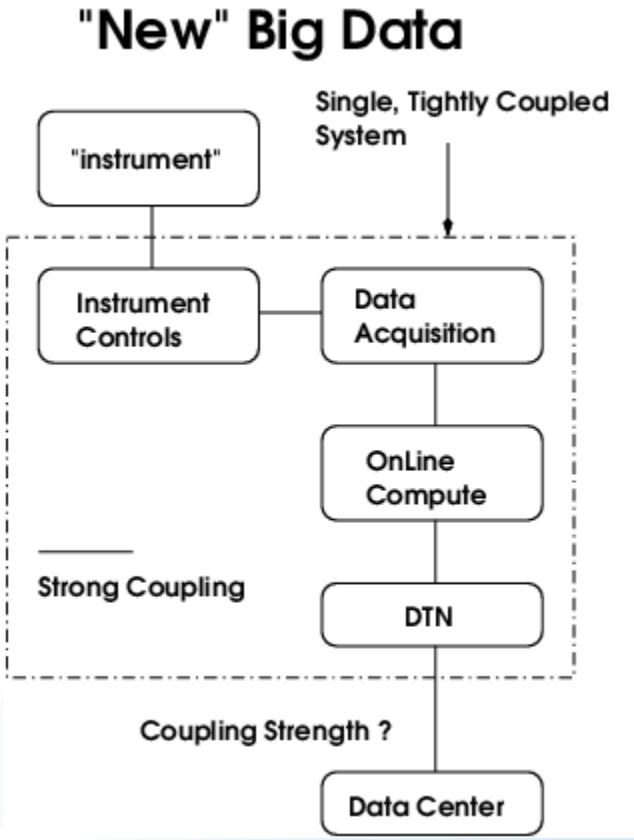
- Achieve “economies of scale” on systems and support
- Location independence of compute and storage
- “Re-taskable” compute and storage. As required
  - Move resources between experiments
  - Move resources between on line and off line
- Enable high bandwidth, low latency movement of data.
- Minimize data movement when/where possible.
- Maintain security of instruments and DAQ systems

Big Data System Architecture



Traditional vs New Big Data





“Simple Solution”

- Single network connecting all sites to the data center has problems
  - Each site may be a separate AA domain
  - Remote site “untrustworthy”
  - NxN inter-site visibility
  - All of the above = Security problem
- Potential fixes to security problem
  - Firewall adds costs, limits performance
  - DTN adds costs and complexity, may limit performance
  - Authenticating when connecting to remote site adds complexity

“Simple Solution” (cont’d)

- Single compute and storage pool has problems
  - On line resources require stronger availability guarantees
    - Potential inter-site interference with shared resources
    - Can off line batch systems accommodate on line requirements ?
  - Integrity and confidentiality of data hard to ensure.
  - “Fast, Cheap, Reliable” choose two for storage.

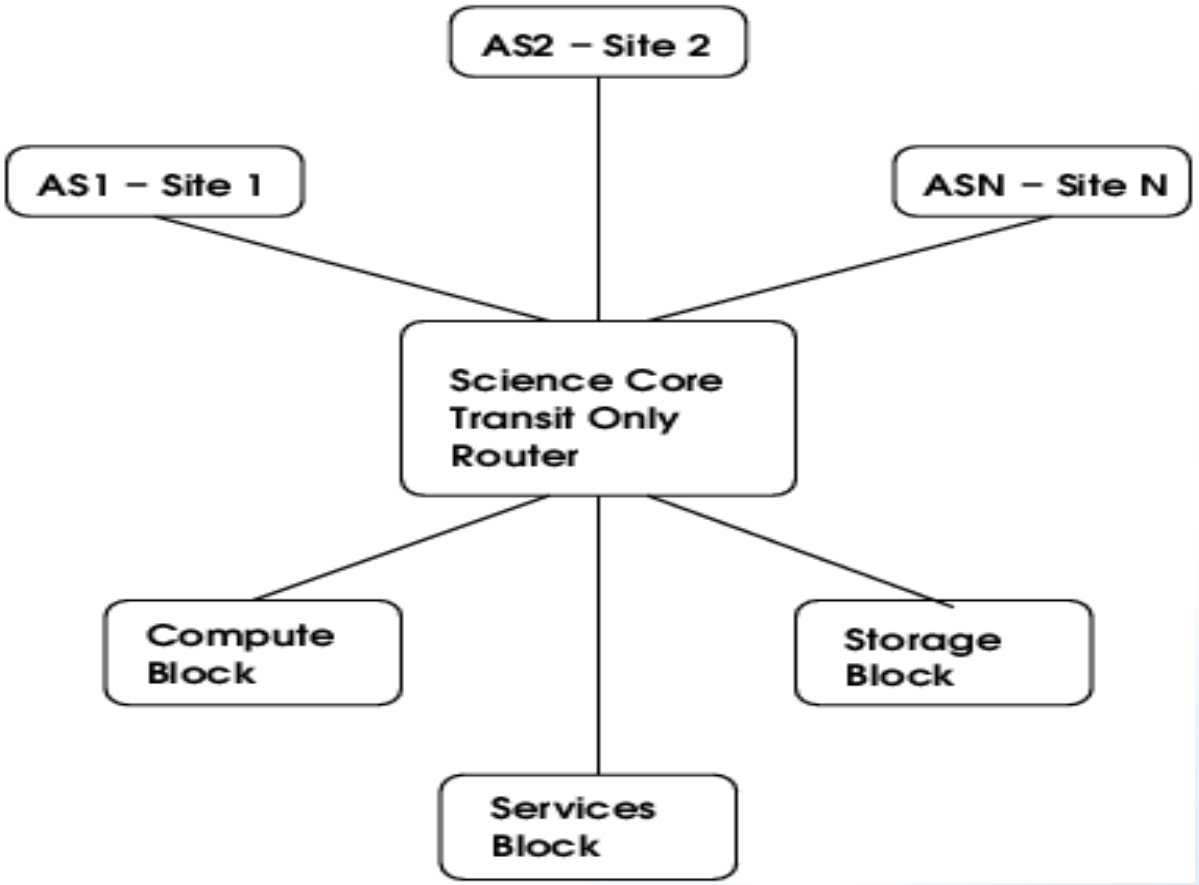
Steps Towards a Solution

- Establish baseline availability, “durability”, and security requirements for compute and storage with instrument sites.
- Establish network connectivity to instrument site while satisfying performance and security requirements
- Design, configure, and deploy the compute and storage infrastructure at the data center to support the requirements

Science Core Network

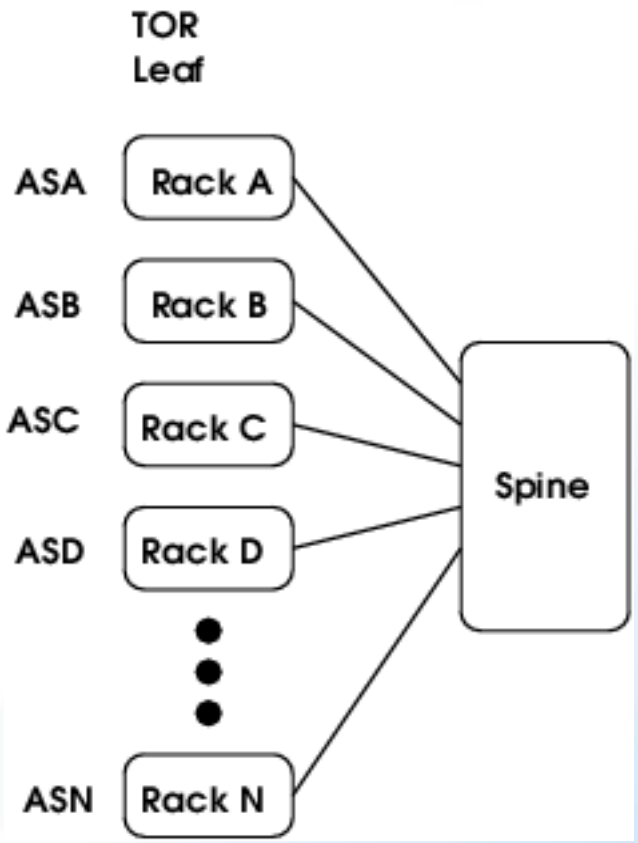
- Network architecture/infrastructure enabler for solving the big science problems.
- Border Gateway Protocol (BGP) is the first of two key linchpins in Science Core architecture
- “Transit only” router core is the other linchpin
- Network partitioned into islands. Specifically, BGP “autonomous system”(AS)
  - Each remote site is an AS
  - Data center services/systems are also selectively placed into autonomous systems

Science Core Network (First Step)



Compute Resources at the Data Center

- Spine+Leaf w/BGP
- Each compute rack is an AS
- Finer granularity possible with overlay networks



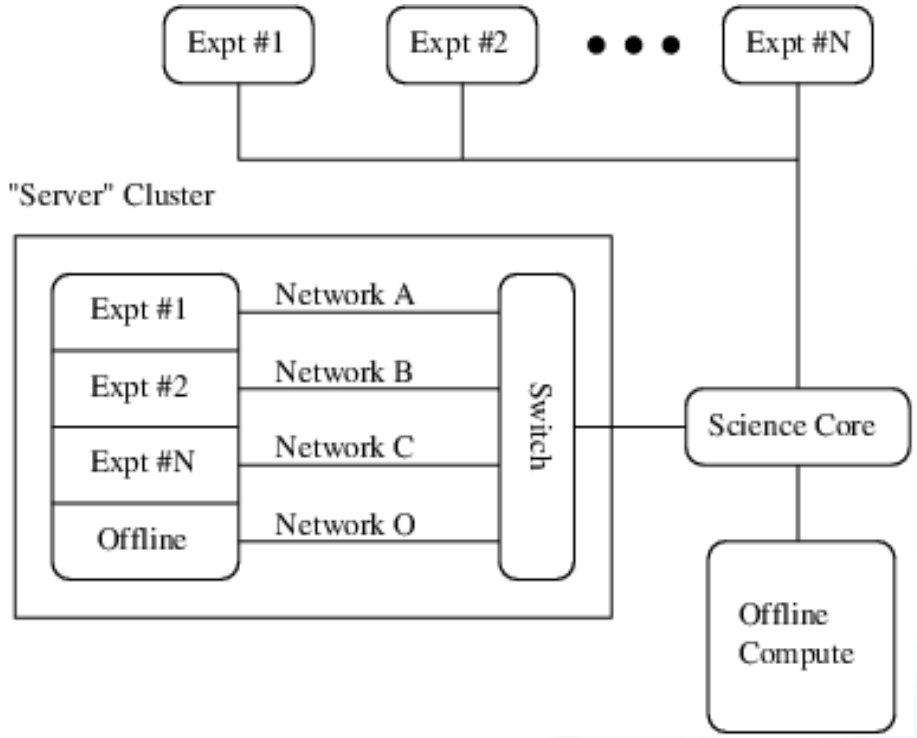
Storage and Services

- Storage and services can be partitioned into multiple AS's
  - More complex network management
  - But additional isolation and security
- Storage and services can also be bundled into fewer AS's for more global visibility
  - Requires secured (hardened) services
  - Sites must trust service administrators
  - Services must not support interactive services
  - Entails additional risks
  - Simpler to administer and understand

Building on the Science Core Network

- Create “off line data center” by enabling visibility between compute, storage, and service AS’s
- Augment on line compute by moving visibility of selected compute AS’s from “off line data center” to target remote AS.
- Augment on line storage by moving visibility of selected AS from “off line data center” to target remote AS.
- Share storage between remote AS and data center by making storage AS visible to both. Note that remote AS and data center are not visible to each other

Dynamic Resource Allocation



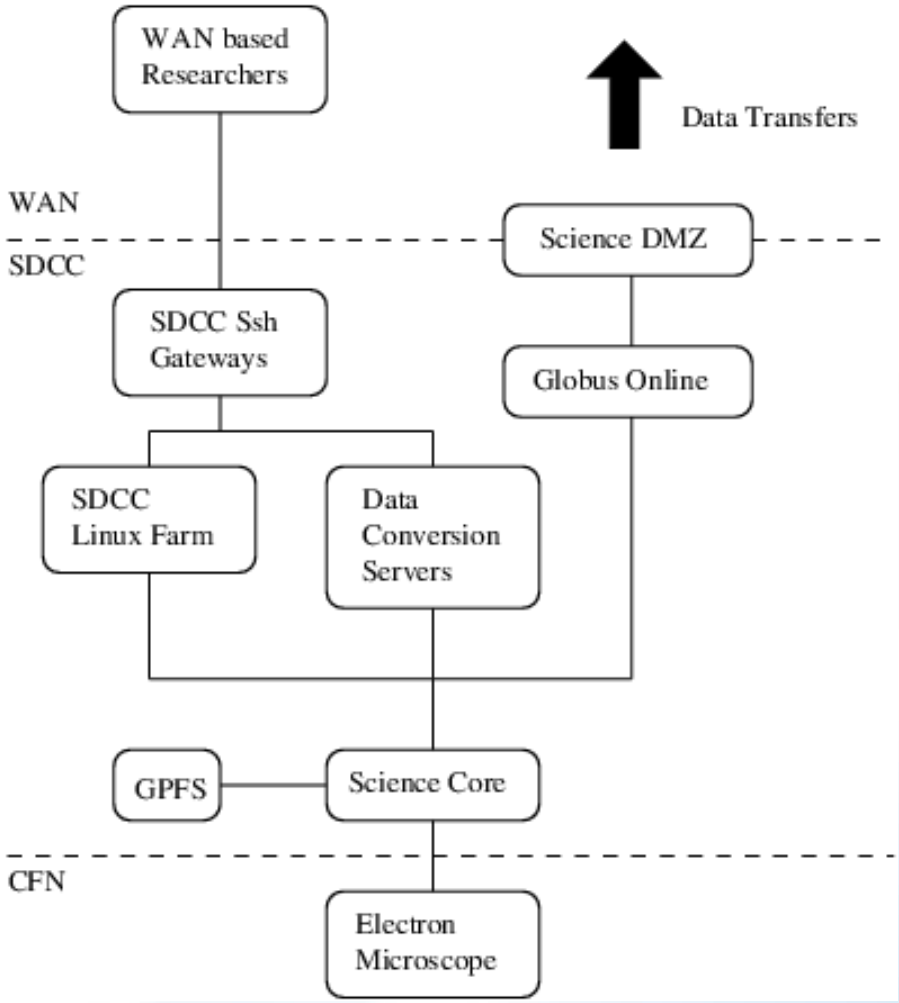
CFN Electron Microscope

- CFN Electron Microscope
  - K2 Direct Detect Imaging System at CFN
  - Generates ~15TB in 5 minutes
  - K2 writes data to internal SSD buffers.
  - Windows PC moves data from SSD to mounted file system
  - As shipped, system was a “data island”
  - GPFS high performance file server at the SDCC

Science Core In Action

- 2x10GbE link established between CFN and SDCC
- Local Windows file system replaced with GPFS file system, resident at the SDCC
- GPFS file system mounted on HPC/HTC and WAN transfer resources
- Note that the K2 system and the Windows PC remain inaccessible from the HPC/HTC and WAN transfer resources at the data center.

CFN Block Diagram



CFN Solution Features

- No software development effort
- Minimal modifications to proprietary K2 DAQ system
- No changes to experiment work flow (except faster turnaround)
- Substantially reduced data readout time (K2 DAQ limited)
- Data “immediately” available for analysis or WAN transfer
- More sophisticated support possible

CFN: Leveraging SDCC Capabilities

- Immediate access to existing HTC resources at the SDCC (on an opportunistic basis only)
- Data immediately available on BNL Institutional HPC cluster at the SDCC
- Data immediately to Globus Online gateways for data transfers out of BNL
- Optional access to SDCC HPSS mass storage system

Open Question from CFN Experience

- Cost can be an open issue
- Are shared storage resources in the data acquisition path a good idea ?
- Can off line compute and on line compute share storage ?
- Will shared compute resources work ?
- Is security a problem when streaming data to production off line compute without “DTN” ?

Big Data and SDCC

- Science Core network is a key enabler for building solutions for the expected deluge of big data experiments.
- Variations in “processing” models is limited only by imagination.
- Creation of processing systems requires input from the remote sites
- Note that the “Laws of Economics” are not changed by Science Core. Long distance, high bandwidth connectivity costs $$. (Still must pay to play)
-

Moving Forward

- Building complete solutions will require close cooperation between multiple groups
  - Researchers at the different instruments
  - DAQ teams at the different instruments
  - Software developers (of all favors)
  - Data center staff (SDCC)
  - WAN (ESNET) and LAN (ITD) networking groups
  - Cybersecurity