ASCR INTELLIGENT OPTICAL NETWORK INFRASTRUCTURE WORKSHOP

(Smart Networks & Middleware Services for Open Distributed Science BigData Infrastructures)

U.S. DEPARTMENT OF ENERGY,

OFFICE OF ADVANCED SCIENTIFIC COMPUTING RESEARCH

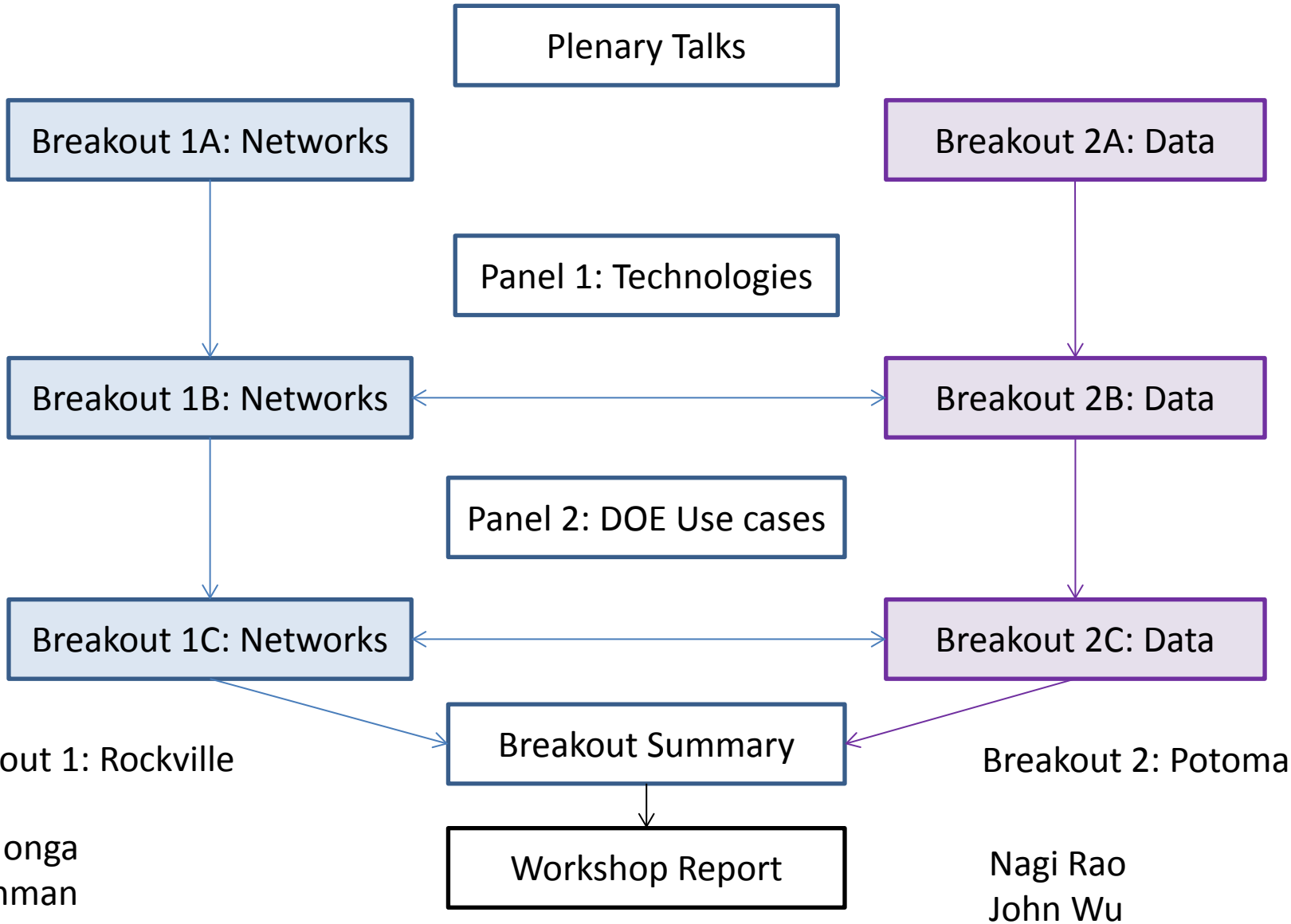
August 5-6, 2014

Gaithersburg, Maryland

U.S. DEPARTMENT OF ENERGY | Office of Science

Intelligent Optical Network Infrastructure Workshop

Organization and Flow



Breakout 1: Rockville

Inder Monga
Tom Lehman

Breakout 2: Potomac

Nagi Rao
John Wu

Intelligent Optical Network Infrastructure Workshop

Outline of Report: (to guide the breakout sessions)

Executive Summary
Background
DOE applications and use cases – 2-3 potentially game changing visions

Current State-of-the-art: emphasis on limitations, impediments high pay-offs
- **Use cases**
- **R&D areas**
- **Next steps for high pay-off technologies and strategies**

Requirements and visions for most impacts on DOE capabilities

Breakout 1A-2A

Technical Tasks and Areas:
- **Networking areas at all levels and SDN areas**
- **Distributed data and storage areas**

Breakout 1B-2B

Recommendations
- **Technical Areas: SDN, storage, file systems**
- **Collaborations: users vendors, facility providers, users and R&D**
- **Multi-agency investments: DOE leverage other agencies and vice versa**

Breakout 1C-2C

Breakout Group Assignments

INS Breakout Only (Total 18)
Jeff Boote (Sandia)
Paul Curtis (Ciena)
Phillip Demar (FNAL)
Andrea Fumagalli (UT Dallas)
Chin Guok (ESnet)
Deniz Gurkan (Univ Houston)
Ron Hutchins (Georgia Tech)
Robert Jacobs (ANL)
Dimitrios Katramatos (BNL)
Greg Lauer (BBN)
Tom Lehman (UMD/MAX)

Park Fields (LANL)
Malathi Veeraghavan (Univ of Virginia)
Venkat Viswanath (ANL)
Vinod Vokkarne (Univ Massachusetts Lowell)
Bing Wang (Univ of Connecticut)
Ben Yoo (UC Davis)

Breakout Group Assignments

FedData Breakout Only (Total 9)
Todd Bowman (LANL)
Garrett Granroth (ORNL)
Anthony Lentine (Sandia)
Peter Nugent (LBL)
Emily Pahlavan (Fermi)
Amedeo Perazzo (SLAC)
Donald Petravick (Univ of Illinois)
Nagi Rao (ORNL)
Wenji Wu (FNAL)
Dantong Yu (BNL)

Breakout Group Assignments

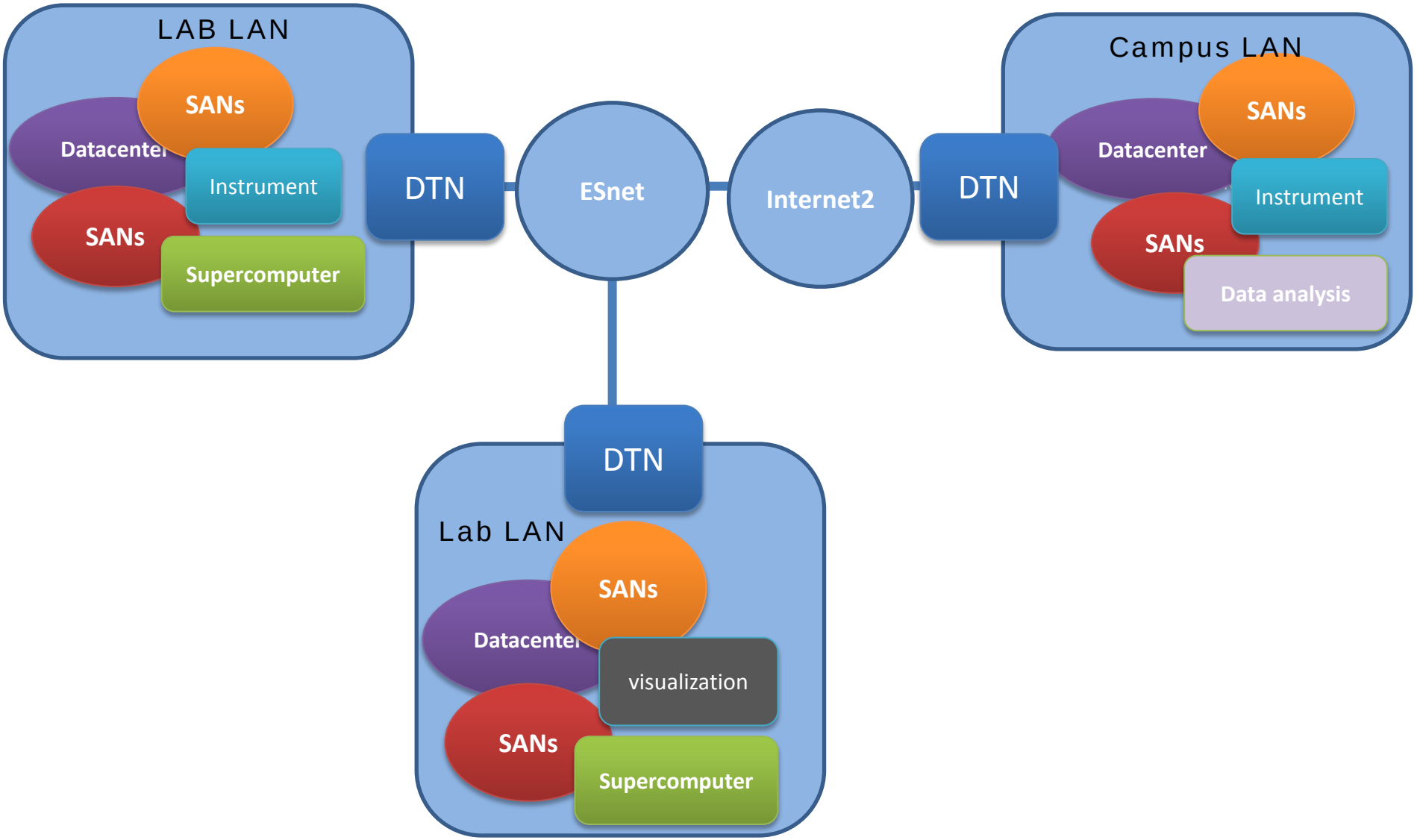
Both INS and FedData Breakout Sessions (Total 9)
Donagh Buckley (EMC)
David Cohen (Intel)
Shu Huang (RENCI)
DK Panda (Ohio State)
Steve Schwab (USC/ISI)
Brian Tierney (ESnet)
Xi Yang (UMD/MAX)
Dean Williams (LLNL)
John Wu (LBL)

Broad Themes

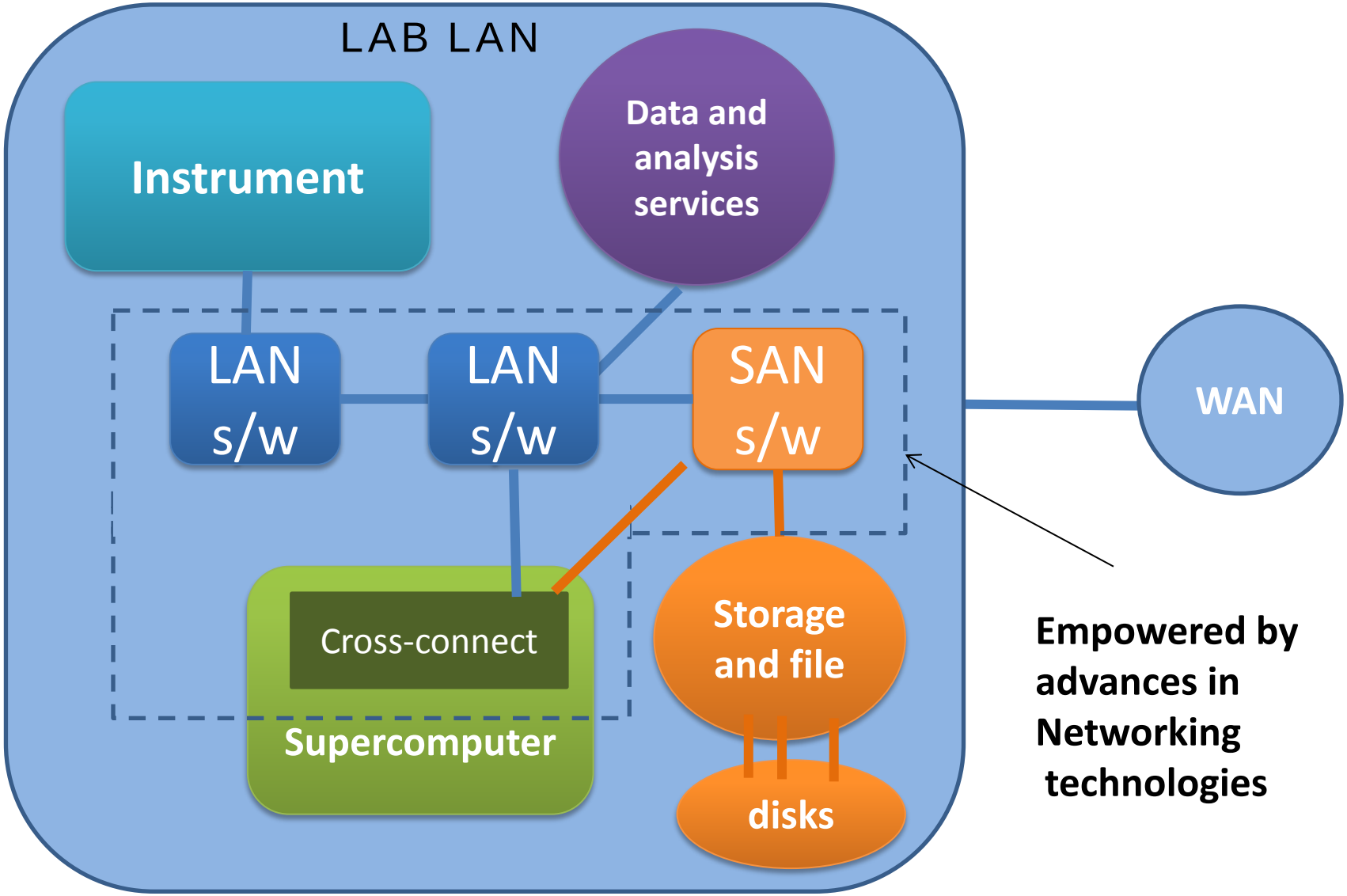
Intelligent Networks, SDN, Distributed Data Infrastructures

- **assess current network technologies** in light of extraordinary advances in computing technologies and the emergence of Big Data applications and workflows
- **identify the challenges and opportunities in transforming current network technologies** to provide intelligent and automated services to accelerate scientific discoveries in the era of distributed extreme-scale computing and pervasive data-centric science
- Identify challenges and research areas associated with **development and deployment of distributed data and storage infrastructures**

E2E SDN-Enabled Internetworking Context



Site Data Services Over SAN/LAN



Technical and Administrative Demarcation Points

- Can we define distinct demarcation points at technical, administrative, and service boundaries?

Network Segment / Services	SANs, DataCenters, LANs	Regional and WANs
SDN-enabled services	SDN-enabled network and middleware services	SDN-enabled Federated Regional and WAN services
Distributed Science BigData	SDN-enabled SANs/LANs services for BigData	SDN-enabled services for distributed BigData

Vision: End-to-End SDN-Oriented Networking
(SANs, Storage Systems, Datacenters, Metadata, File-systems)

SDN, SANs, Storage Systems, BigData		
Areas/Topics of Interest	SDN-Enabled Network Services	SDN-Enabled Data Services
Bandwidth mismatch between SSD memory, hosts and LAN technologies (FC, GigE, IB)		
SANs traffic engineering with SDN? Performance and security		
Consolidation of SCSSI, FC, IB, Ether, etc. in common protocol framework		
Disks-to-network I/O		
FSIO – Metadata – data format		
Storage virtualization/cloud		
Data server node: QoS, provisioning; multi-core, multi-bus hosts		
Data movement: TCP/UDP/others, data movement services		
Datacenter networking: traffic engineering, energy, security		
Supercomputer data transfer and steering		
Instrument/experimental facilities data and control: SDN reach and support		
Security and vulnerabilities: host, SAN, LAN, data center, WAN		
Operations and performance: technology maturity and performance; transition paths		

Breakout Material

Intelligent Network Infrastructure and Services

- What are the “Services” that next generation distributed BigData Applications and workflows need from networks (SAN/LAN/MAN/WAN)?
- What are the specific services that distributed data infrastructure will need?
- Is it all about End-to-End QoS which spans network regions and storage systems?
- Are Network Virtualization and application specific topologies important?

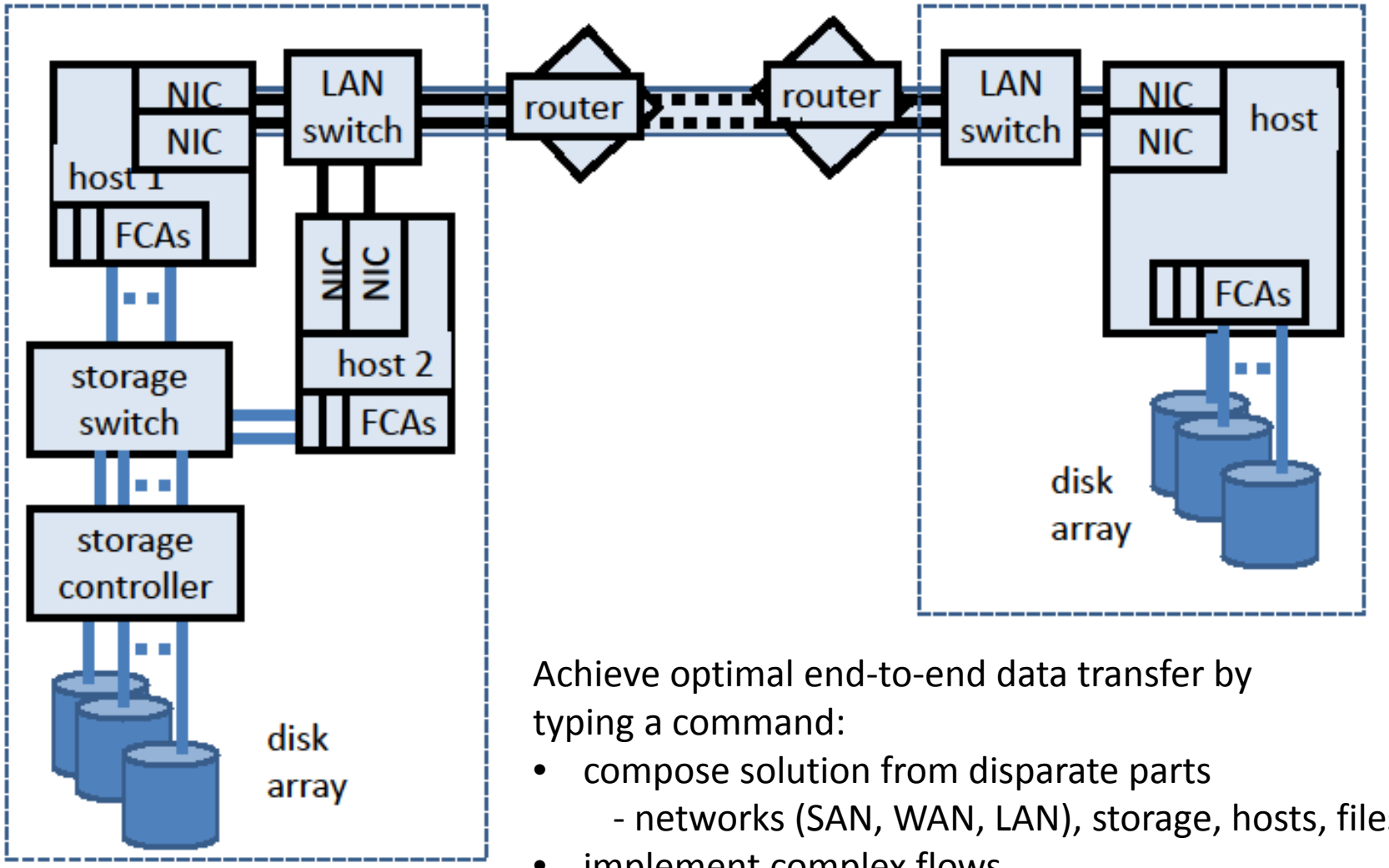
Software Defined Networks (SDN)

- What is the objective of SDN?
  - To enable a more flexible network so that operators can develop “Intelligent Services” for their users and manage their own network more easily?
  - To give “users” more direct access and control over network resources?
  - Is it all about Network Virtualization? Or FlowSpace?
- Where does DOE use case fit with respect to the paradigms of XaaS (Anything as a Service)
  - IaaS, PaaS, NaaS, SaaS, MaaS (Infrastructure, Platform, Network, Software, Monitoring…as a Service)

Distributed Data Infrastructures

- What are the key services that distributed storage and data infrastructures will provide to users?
- What are the key services that distributed storage and data infrastructures will need from next generation networks?

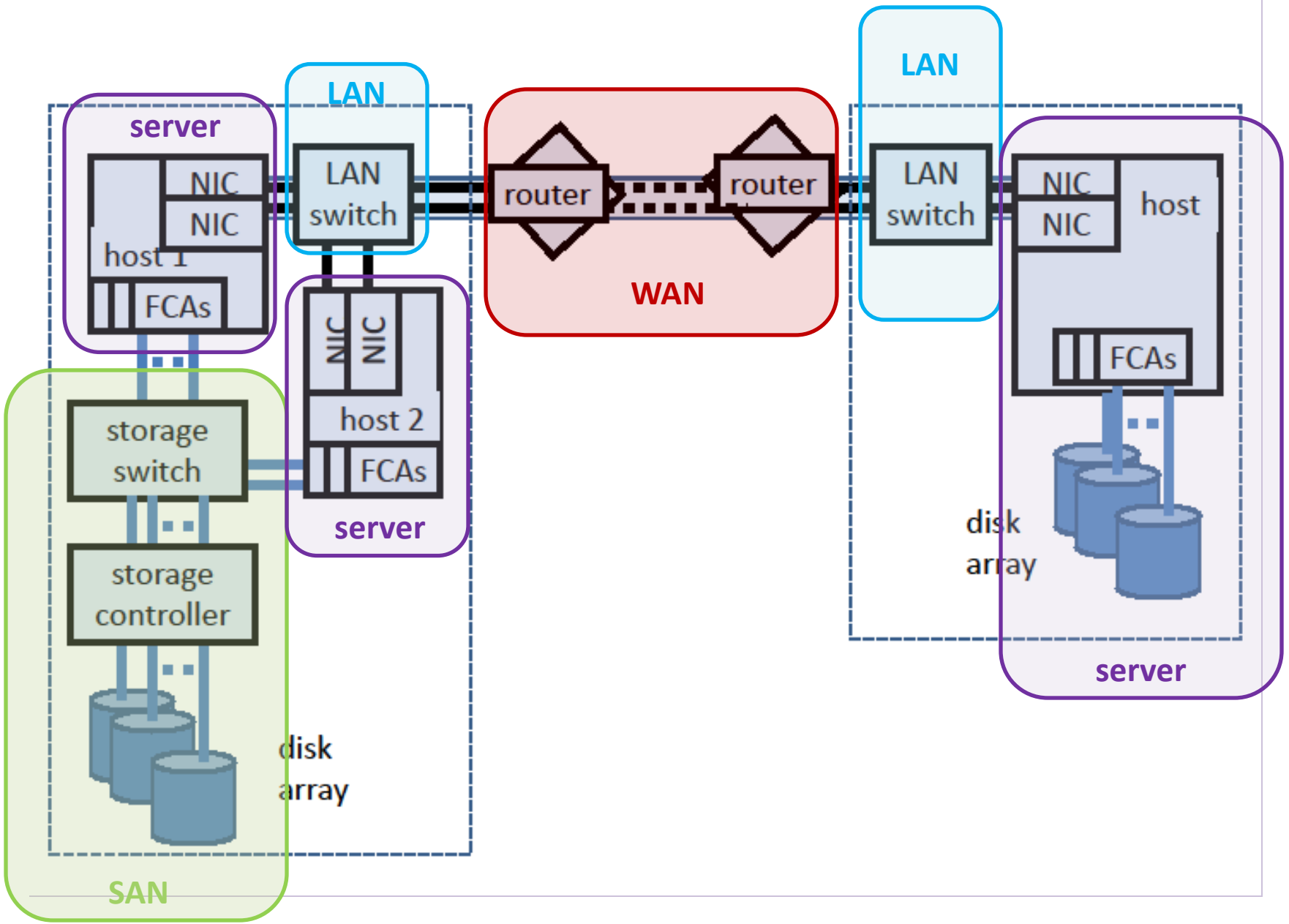
Multiple Multi-Bus Host-to-Host Disk/File Transfers



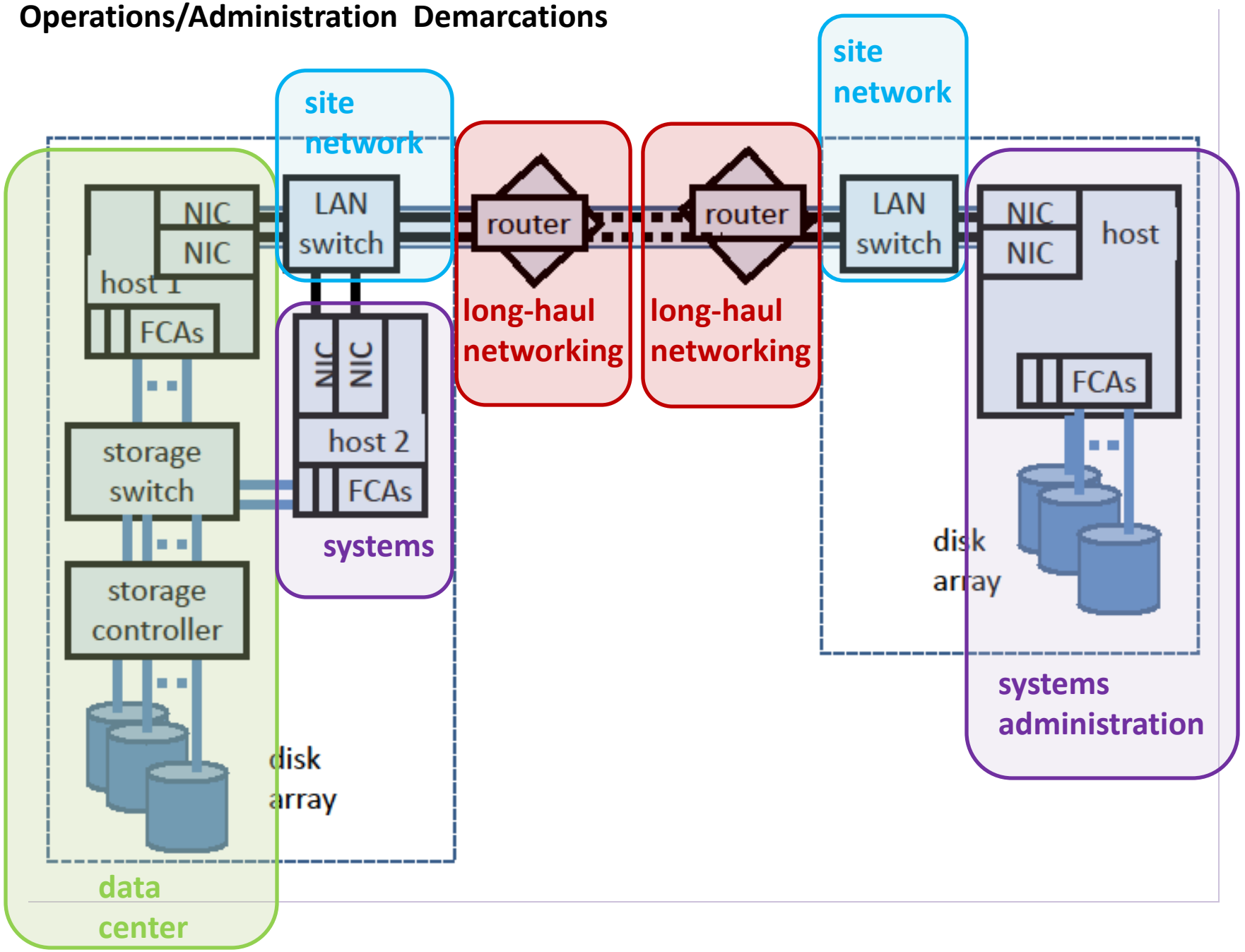
Achieve optimal end-to-end data transfer by typing a command:

- compose solution from disparate parts
  - networks (SAN, WAN, LAN), storage, hosts, files, …
- implement complex flows
  - set-up, configure, impedance match, monitor, …

Technology Demarcations



Operations/Administration Demarcations



General Guidelines Breakout 1A
Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking

Current state of the art, missing capabilities:

- Identify current networking and middleware technologies available and in use for supporting big data applications
- Identify current limitations and impediments with respect to supporting current and emerging big data science applications?
- Identify 2-3 specific services or capabilities that next generation networks should be able to provide to better support distributed big data applications and workflows
- Consider these questions as applied to wide area, regional, campus/laboratory/enterprise, and datacenter infrastructures.

General Guidelines Breakout 1B

Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking

Technical Tasks and Areas:

- Identify technical areas and challenges in intelligent network services and middleware systems
  - network provisioning and topology management
  - Network virtualization
  - Software Defined Networking (SDN)
  - OpenFlow and related technologies
  - How to design, test, deploy and maintain services
- How can intelligent network services and SDN provide benefits to distributed data and storage system infrastructures and big data application workflows?
  - what types of network and middleware services are needed to support emerging big data driven applications and use cases?
  - network embedded storage and/or compute resources?
  - Software Defined Infrastructure (SDI) role?
- What are the specific, high-impact science data capabilities empowered by Intelligent SDN and related technologies?
- What are the proper demarcations points and services between SAN, LAN, ScienceDMZ, and WAN?
- What are data security implications and technologies needed?

General Guidelines Breakout 1C
Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking

Recommendations:

- Identify priority areas and develop recommendations, including:
  - technologies
  - operations and provisioning
  - security and vulnerability considerations
  - Programmatics
  - multi-agency collaborations

General Guidelines Breakout 2A

Federated Science BigData infrastructures with SDN-enabled Services and Workflows

Broader Strategic Session Goals:
Identify current data and storage technologies for supporting big data, science applications,

- emphasis on DOE use cases

What are current limitations and impediments in reaping full potential of these services?
Identify 2-3 specific cases that can achieve next generation capabilities and performances with advances in data and storage systems.

Specific Technical Areas: Enabled by advances in SDNs and Networking:

Integration of LANs and WANs provisioning and control into data applications
SAN technologies that control storage and networking devices
Data and analysis servers with certain control on LAN, SAN and WAN
File systems enabled by “flow controls” across data center, supercomputer
Instrument control with big and control flows across networks
Performance, functionality, technology maturity
Operations support, phase-in mechanisms and strategies
Security and vulnerability

General Guidelines Breakout 2B

Federated Science BigData infrastructures with SDN-enabled Services and Workflows

Technical Tasks and Areas:

Identify technical areas in data and storage systems that cross-fertilize with network services for improved capabilities and performance

- data and storage services
- file systems and metadata services
- design, test, deploy and maintain services

How can data and storage services benefit from programmable networks and related services?

- SDN SANS
- host-programmable SANs, LANs, WANs

What are the specific, high-impact science data capabilities empowered by SDN and related technologies?

What are data security implications?

- intentional and incidental effects

General Guidelines Breakout 2C

Federated Science BigData infrastructures with SDN-enabled Services and Workflows

Identify priority areas and develop recommendations, including:

- technologies
- operations and provisioning
- security and vulnerability considerations
- programmatics
- multi-agency collaborations

Specific Technical Areas: Enabled by advances in SDNs and Networking:

- Integration of LANs and WANs provisioning and control into data applications
- SAN technologies that control storage and networking devices
- Data and analysis servers with certain control on LAN, SAN and WAN
- File systems enabled by “flow controls” across data center, supercomputer
- Instrument control with big and control flows across networks
- Performance, functionality, technology maturity
- Operations support, phase-in mechanisms and strategies
- Security and vulnerability

General Guidelines for Panelists

What are state-of –the –art or current capabilities?
What are the requirements and desired capabilities?
What are expected evolutionary trajectories and capabilities not addressed?
What are the needed R&D areas and tasks for quantum leaps in capabilities?
What are the needed programmatic and cross-fertilization tasks?
What are the operations and deployment considerations?
What are the vulnerabilities and security considerations?
What are top 3-5 important focus areas/tasks in areas of your expertise?