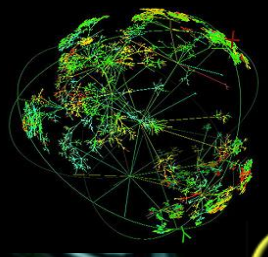
SC20 Advanced Networking For Data Intensive Science Experiments and Demonstrations

Joe Mambretti, Director, (j-mambretti@northwestern.edu)
International Center for Advanced Internet Research (www.icair.org)
Northwestern University
Director, Metropolitan Research and Education Network (www.mren.org)
Co-Director, StarLight (www.startap.net/starlight), Director, StarLight International/National Communications Exchange Facility (www.startap.net/starlight),
PI IRNC: RXP: StarLight SDX, Co-PI Chameleon, PI-iGENI, PI-OMNINet

JET Meeting
November 17, 2020
International Conference On High Performance Computing Networking, Storage and Analytics
November 9-20, 2020



Selected Applications

GENI
www.geni.net

Open Storage Network
www.openstoragenetwork.org

OSG
www.opensciencegrid.org

Polar Geospatial Center
www.pgc.umn.edu

LSST
www.lsst.org

Virgo
www.virgo-gw.eu

LHC
home.cern/science/accelerators/large-hadron-collider

GLEON
www.gleon.org

OSiRIS
www.osris.org

GRP
theglobalresearchplatform.net/

IceCube
icecube.wisc.edu

Pierre Auger Observatory
www.auger.org

ISS
www.nasa.gov/station

LIGO
www.ligo.caltech.edu

XSEDE
www.xsede.org

PRP
pacificresearchplatform.org

Chameleon
www.chameleoncloud.org

SKA
www.skatelescope.org

LHCONE
twiki.cern.ch/twiki/bin/view/LHCONE/WebHome

USGS EROS
www.usgs.gov/centers/eros

Blue Waters
bluewaters.ncsa.illinois.edu

Jetstream
www.jetstream-cloud.org

Belle II
www.belle2.org

XENON
xenon.astro.columbia.edu

SDSS
www.sdss.org

LHCOPN
twiki.cern.ch/twiki/bin/view/LHCOPN/WebHome

PRAGMA
www.pragma-grid.net

CHASE-CI
www.calit2.net/newsroom/article.php?id=2910

NEON
www.neonscience.org

CENTRA
www.globalcentra.org

SAGE2
sage2.sagecommons.org

Genomic Science Program
genomicscience.energy.gov

LBNF/DUNE/ProtoDUNE
lbnf.fnal.gov

NOvA
novaexperiment.fnal.gov

ALMA
www.almaobservatory.org

IVOA
www.ivoa.net

Compilation by Maxine Brown and Joe Mambretti

STARLIGHT℠

GRP: Services, Architecture, Technology

- Architecture: “Global Science DMZ”
- Services Optimized For Science Workflows
- High Performance Transport Over WANs For Large Capacity Data Streams E2E
- Enhanced Virtualization
- Highly Programmable
- Specialized Components (Software Stacks, Next Gen DTNs, Advanced APIs, Customized Devices, Federation, INT, P4, etc.)
- Interdomain Dynamic Provisioning
- Production Resources Adjacent to Testbeds

Next Generation Distributed Enviroment For Global Science

STARLIGHT℠

INTERNATIONAL PEERING EXCHANGE
Global Research Platform Network
100 Gbps
StarLight
Global Research Platform

Global Research Platform: Global Lambda Integrated Facility Available Advanced Network Resources

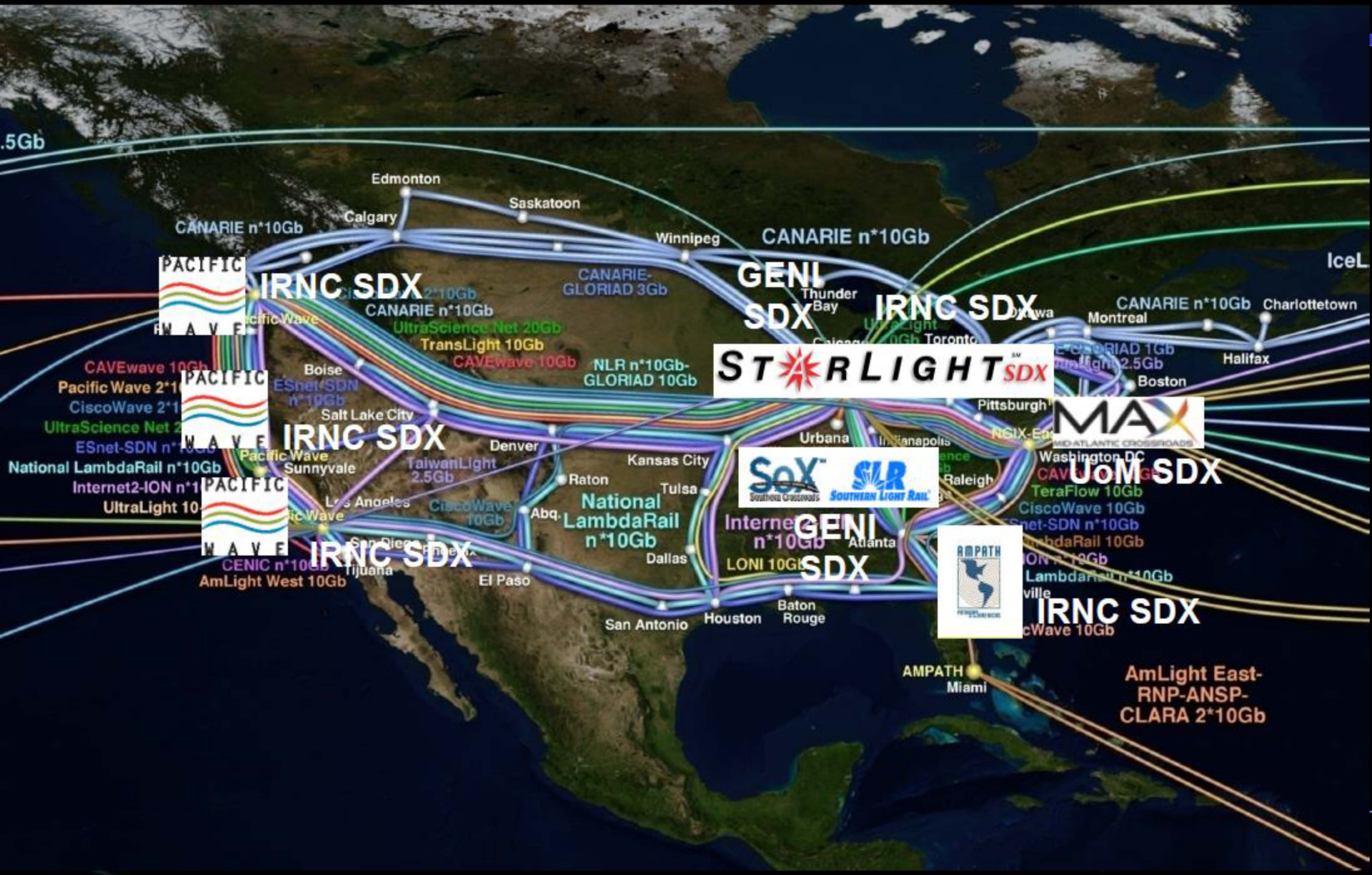
Visualization courtesy of Bob Patterson, NCSA; data compilation by Maxine Brown, UIC.

iCAIR

www.glif.is

STARLIGHT℠

Emerging US SDX Interoperable Fabric



AutoGOLE Fabric 2020

StarLight – "By Researchers For Researchers"

StarLight: Experimental Optical Infrastructure/Proving Ground For Next Gen Network Services

Optimized for High Performance Data Intensive Science

Multiple 100 Gbps (60+ Paths)

StarWave

100 G Exchange

World's Most Advanced Exchange

Multiple First of a Kind Services and Capabilities

View from StarLight

Abbott Hall, Northwestern University's Chicago Campus

iCAIR

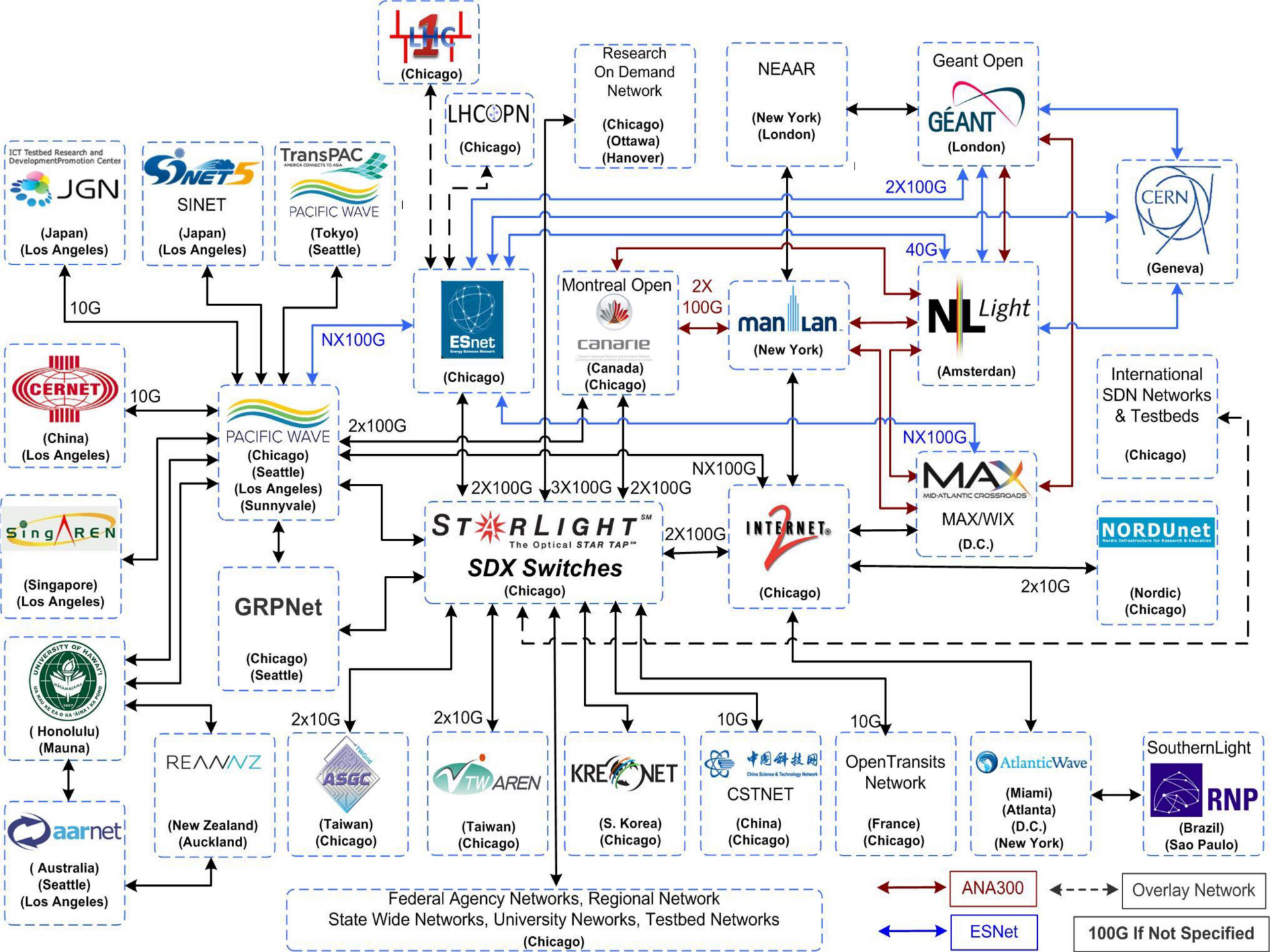
STARLIGHT℠

StarLight SDX Overview

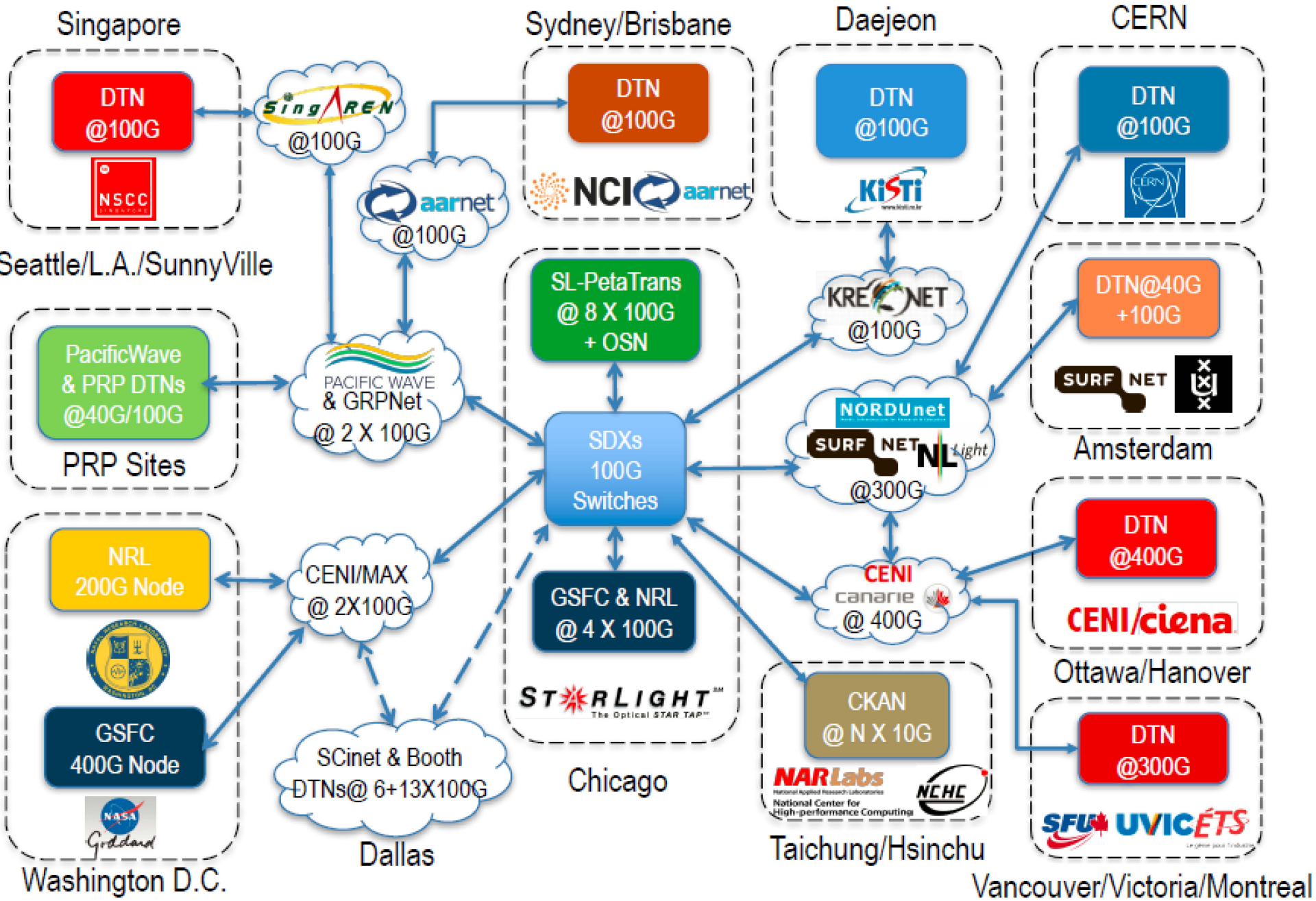
- This IRNC StarLight SDX Initiative Is Designing, Implementing, and Operating New Services For Global Data Intensive Sciences, Based On Emerging Next Generation Architecture and Technolgies, Including Virtualization, Orchestration, Segmentation (Slicing), Software Defined Resources, Programmability and Customization.
- These Macro Trends Enable Exchanges To Be Agile Platforms For Dynamic Services Provisioning, Real-Time Responsiveness, and Distributed Control Over Core Resources, Including By Applications, Edge Processes and Devices.

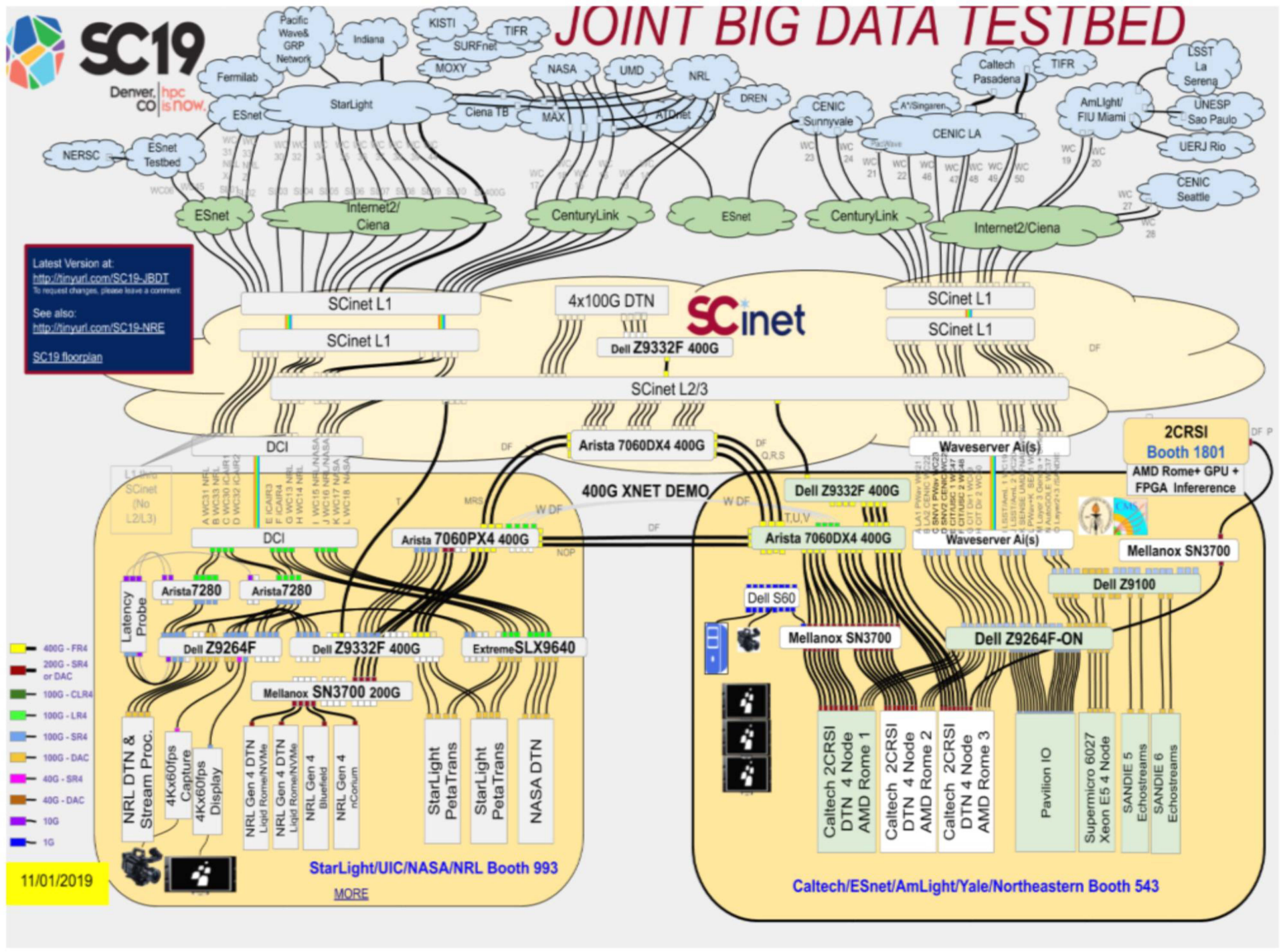
This Project Is Transitioning Network Exchanges To

Open Innovation Platforms

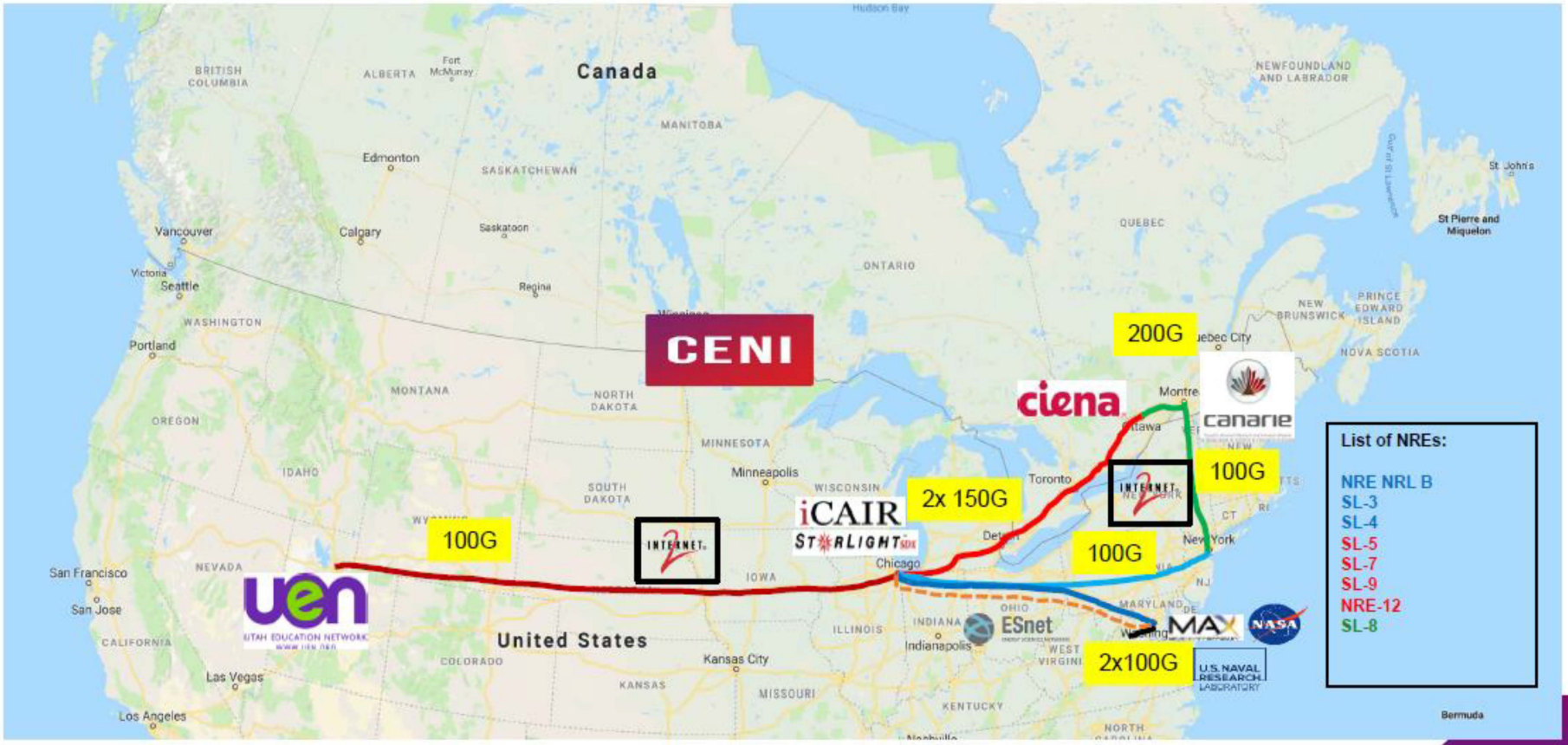

LHC
(Chicago)
LHCOPN
(Chicago)
Research On Demand Network
(Chicago)
(Ottawa)
(Hanover)
NEAAR
(New York)
(London)
Geant Open
GÉANT
(London)
ICT Testbed Research and DevelopmentPromotion Center
JGN
(Japan)
(Los Angeles)
SINET
(Japan)
(Los Angeles)
TransPAC
PACIFIC WAVE
(Tokyo)
(Seattle)
2X100G
CERN
(Geneva)
40G
10G
NX100G
ESnet
(Chicago)
Montreal Open
canarie
(Canada)
(Chicago)
2X
100G
manLan
(New York)
NL Light
(Amsterdan)
CERNET
(China)
(Los Angeles)
10G
PACIFIC WAVE
(Chicago)
(Seattle)
(Los Angeles)
(Sunnyvale)
2x100G
International SDN Networks & Testbeds
(Chicago)
NX100G
NX100G
MAX
MID-ATLANTIC CROSSROADS
MAX/WIX
(D.C.)
2X100G
3X100G
2X100G
SingAREN
(Singapore)
(Los Angeles)
STARLIGHT
The Optical STAR TAP
SDX Switches
(Chicago)
2X100G
INTERNET2
(Chicago)
NORDUnet
(Nordic)
(Chicago)
2x10G
GRPNet
(Chicago)
(Seattle)
(Honolulu)
(Mauna)
2x10G
2x10G
10G
10G
REANNZ
(New Zealand)
(Auckland)
ASGC
(Taiwan)
(Chicago)
TWAREN
(Taiwan)
(Chicago)
KREONET
(S. Korea)
(Chicago)
CSTNET
(China)
(Chicago)
OpenTransits Network
(France)
(Chicago)
AtlanticWave
(Miami)
(Atlanta)
(D.C.)
(New York)
SouthernLight
RNP
(Brazil)
(Sao Paulo)
aarnet
(Australia)
(Seattle)
(Los Angeles)
Federal Agency Networks, Regional Network
State Wide Networks, University Neworks, Testbed Networks
(Chicago)
ANA300
Overlay Network
ESNet
100G If Not Specified

PetaTrans: Petascale Sciences Data Transfer




JOINT BIG DATA TESTBED
SC19
Denver, CO
hpc is now.
Pacific Wave& GRP Network
Indiana
KISTI
TIFR
SURFnet
MOXY
Fermilab
ESnet
StarLight
NASA
UMD
NRL
DREN
Ciena TB
MAX
AtlOnet
Caltech Pasadena
TIFR
LSST La Serena
CENIC Sunnyvale
A*/Singaren
CENIC LA
AmLight/ FIU Miami
UNESP Sao Paulo
UERJ Rio
NERSC
ESnet Testbed
CENIC Seattle
ESnet
Internet2/ Ciena
CenturyLink
ESnet
CenturyLink
Internet2/Ciena
Latest Version at:
http://tinyurl.com/SC19-JBDT
To request changes, please leave a comment
See also:
http://tinyurl.com/SC19-NRE
SC19 floorplan
SCinet L1
SCinet L1
4x100G DTN
SCinet
Dell Z9332F 400G
SCinet L1
SCinet L1
SCinet L2/3
DCI
Arista 7060DX4 400G
Waveserver Ai(s)
2CRSI
Booth 1801
AMD Rome+ GPU + FPGA Infererence
400G XNET DEMO
Dell Z9332F 400G
DCI
Arista 7060PX4 400G
Arista 7060DX4 400G
Waveserver Ai(s)
Mellanox SN3700
Latency Probe
Arista7280
Arista7280
Dell Z9100
Dell S60
Dell Z9264F
Dell Z9332F 400G
ExtremeSLX9640
Mellanox SN3700
Dell Z9264F-ON
Mellanox SN3700 200G
NRL DTN & Stream Proc.
4Kx60fps Capture
4Kx60fps Display
NRL Gen 4 DTN Liqid Rome/NVMe
NRL Gen 4 DTN Liqid Rome/NVMe
NRL Gen 4 Bluefield
NRL Gen 4 nCorium
StarLight PetaTrans
StarLight PetaTrans
NASA DTN
Caltech 2CRSI DTN 4 Node AMD Rome 1
Caltech 2CRSI DTN 4 Node AMD Rome 2
Caltech 2CRSI DTN 4 Node AMD Rome 3
Pavilion IO
Supermicro 6027 Xeon E5 4 Node
SANDIE 5 Echostreams
SANDIE 6 Echostreams
400G - FR4
200G - SR4 or DAC
100G - CLR4
100G - LR4
100G - SR4
100G - DAC
40G - SR4
40G - DAC
10G
1G
11/01/2019
StarLight/UIC/NASA/NRL Booth 993
MORE
Caltech/ESnet/AmLight/Yale/Northeastern Booth 543

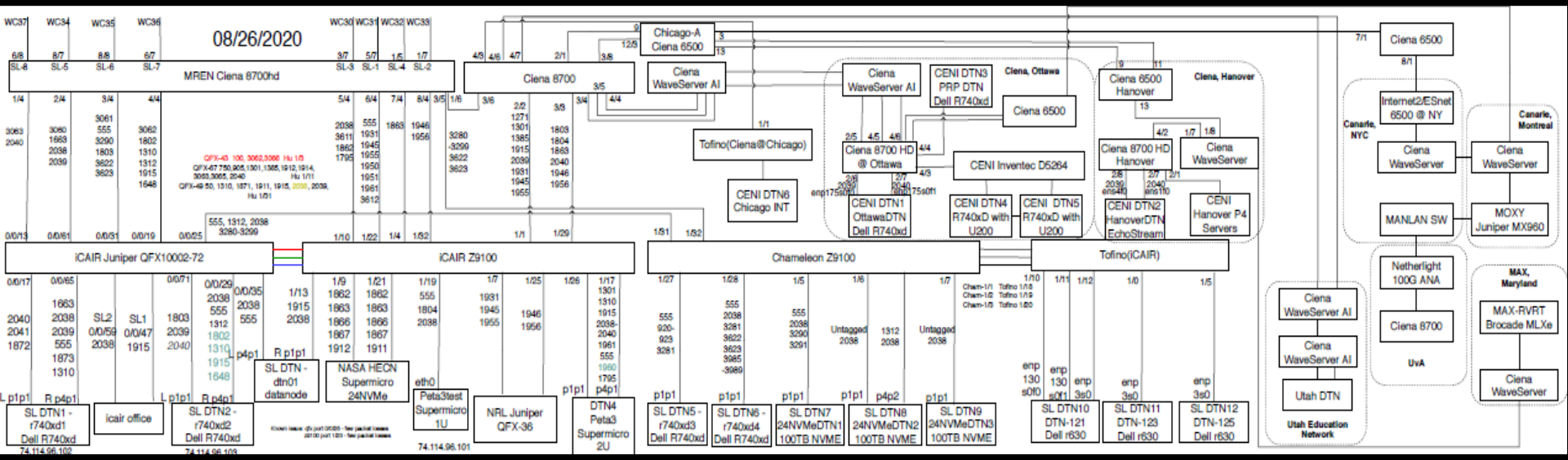
CENI NRE Map for SC20



ciena. Copyright © Ciena Corporation 2017. All rights reserved.

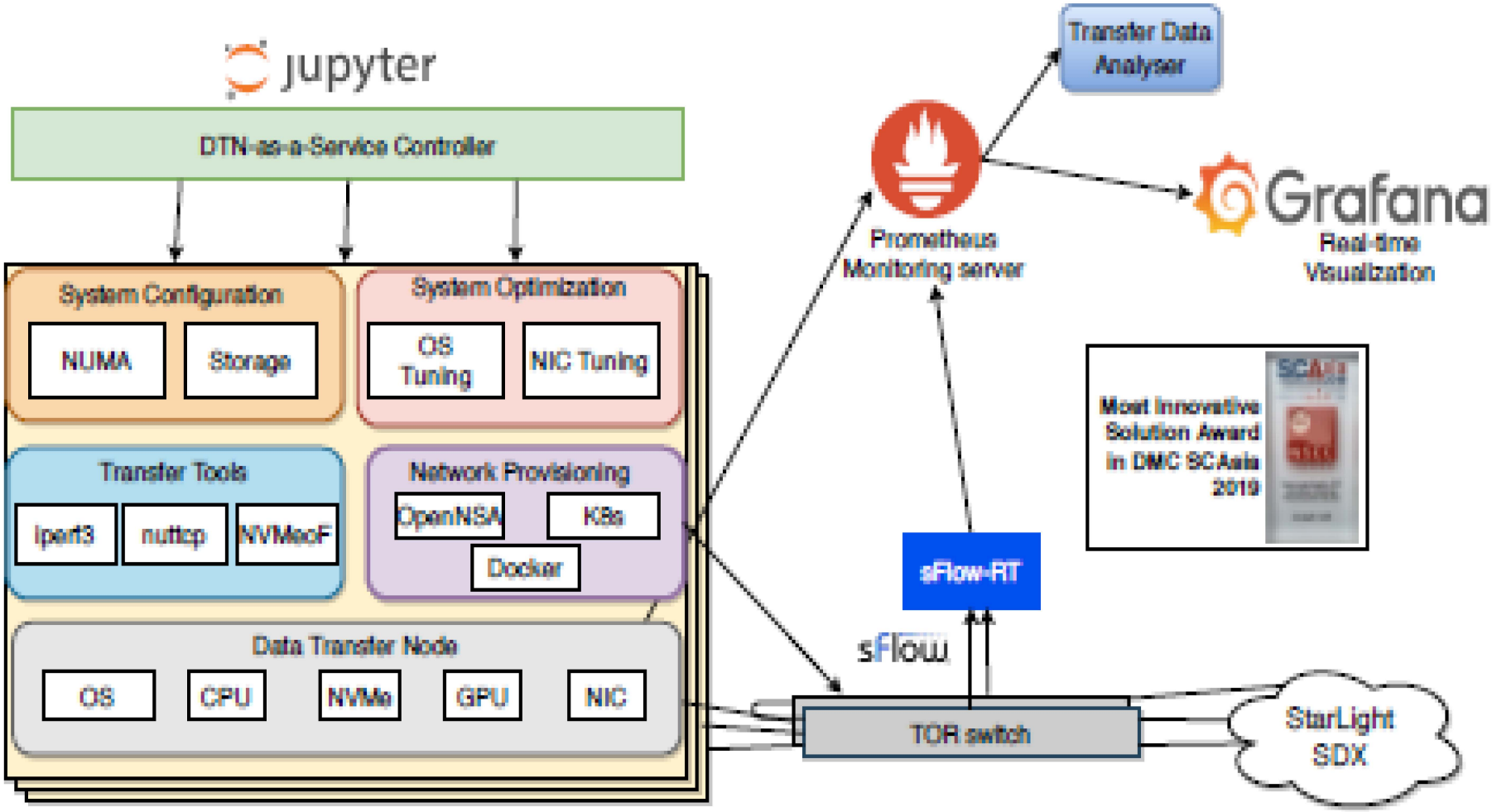
STARLIGHT℠

SC20 Experiments/Demonstrations Testbed

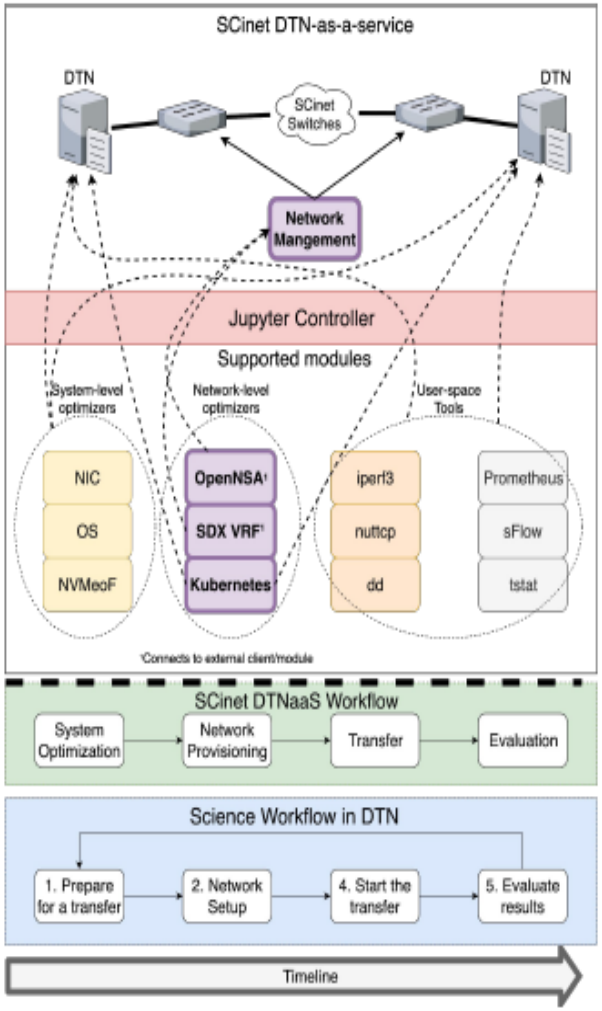


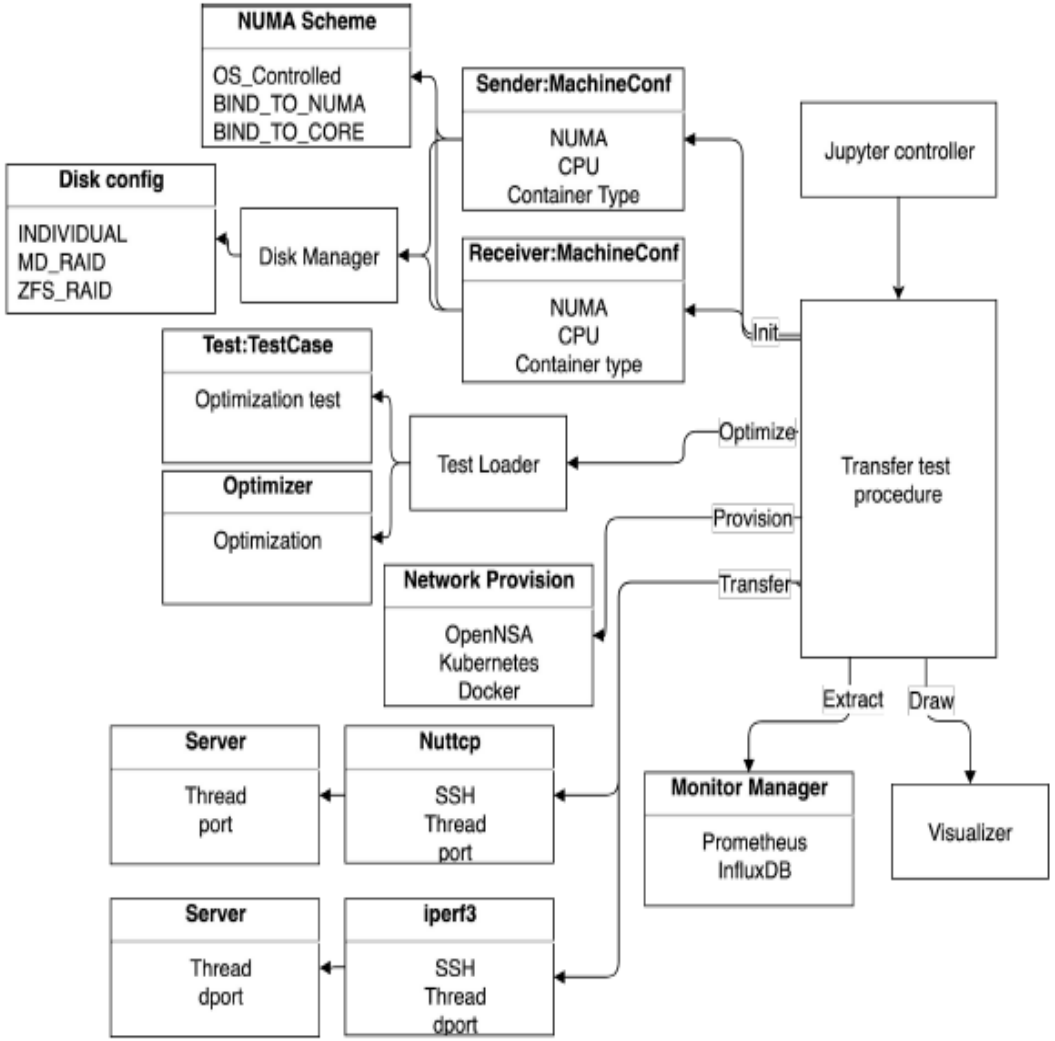
STARLIGHT℠

STARLIGHT SDX DTN-as-a-Service



Design and implementation





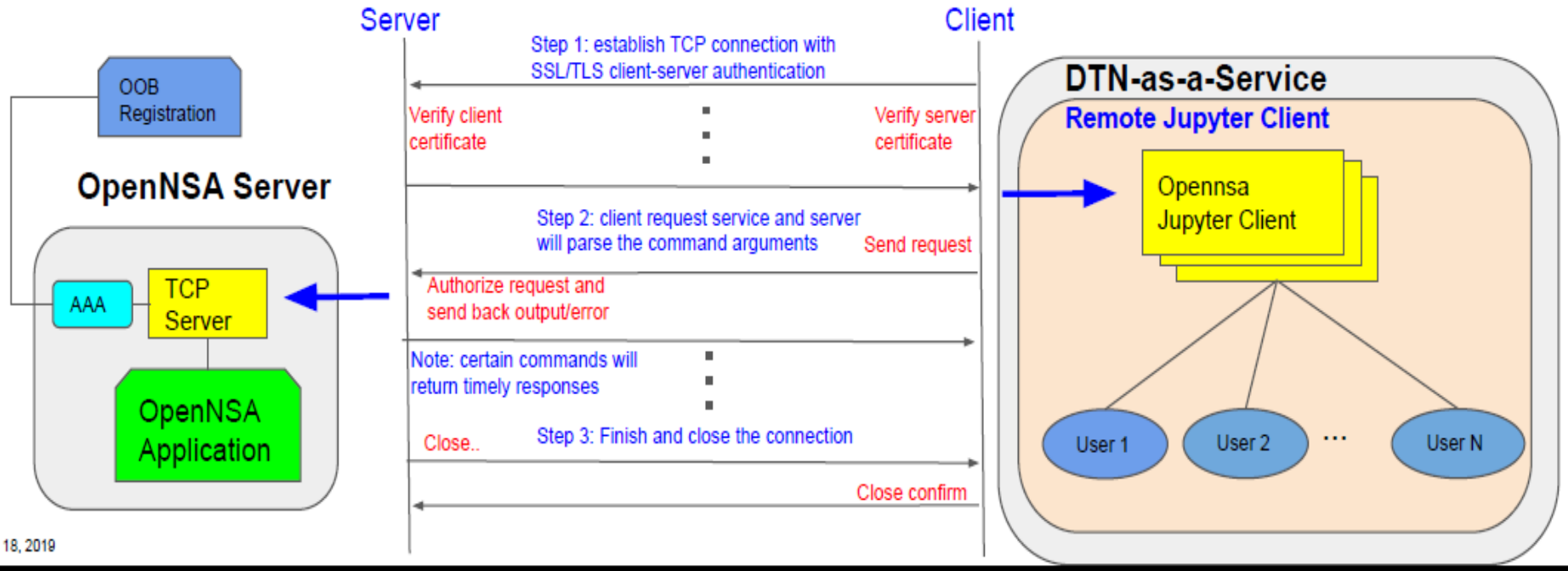
Source: Se Young Yu, iCAIR

STARLIGHT℠

Jupyter Client for NSI OpenNSA Integration

Securely allow users to run NSI OpenNSA services(i.e. dynamically stitch layer 2/3 circuits based on technologies e.g. VLAN.)

- Features: **Authentication, Authorization, Accounting**
- **Authentication**: SSL/TLS operation authenticate both server and client. (certificates need to be exchanged prior unless using public certificates)
- **Authorization**: server will allow access users to request services on certain ports/VLANs based on user identification. I.e. Request command arguments will be parsed and authorized if allowed.
- **Accounting**: for future requirement.
- Additional Feature:Asynchronous(non-blocking operation allows multiple users to request services simultaneously)



Jan 18, 2019

Source: Se Young Yu, iCAIR

STARLIGHT℠

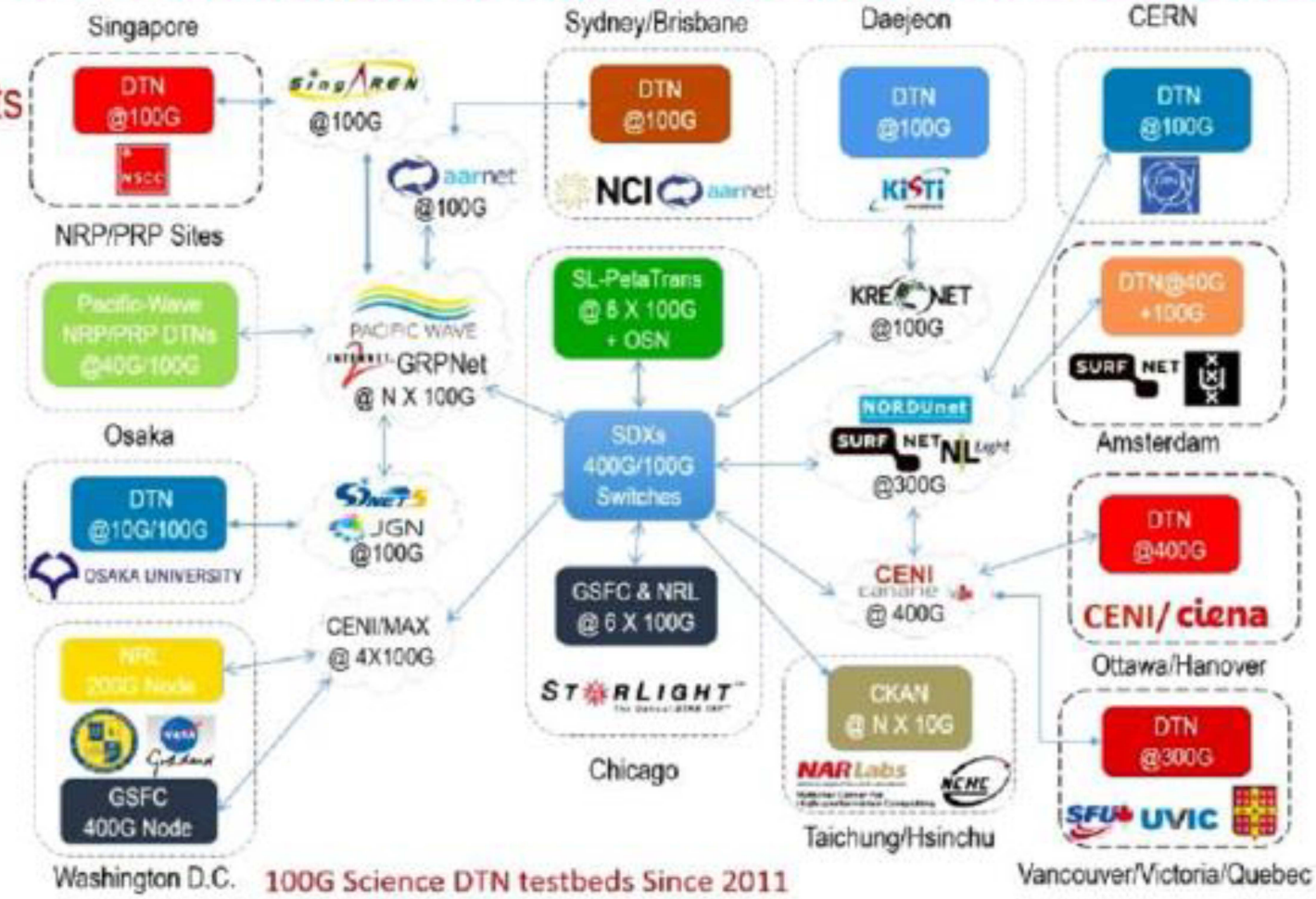
SCinet DTN-as-a-Service @ SC20

- For SC20, DaaS supports XNET Experiments, 12+ NREs: ROCE over WAN, P4 Experiments/Demonstrations, SENSE/openNSA integration, Kubernetes Federation, PCI-e Gen4 DTNs and Many More, Including The “Bring-Your-Own-Testbed Concept”

NRE09-GRP Service: DTNaaS for Petascale Sciences Data Transfer

As 10/09 Supports

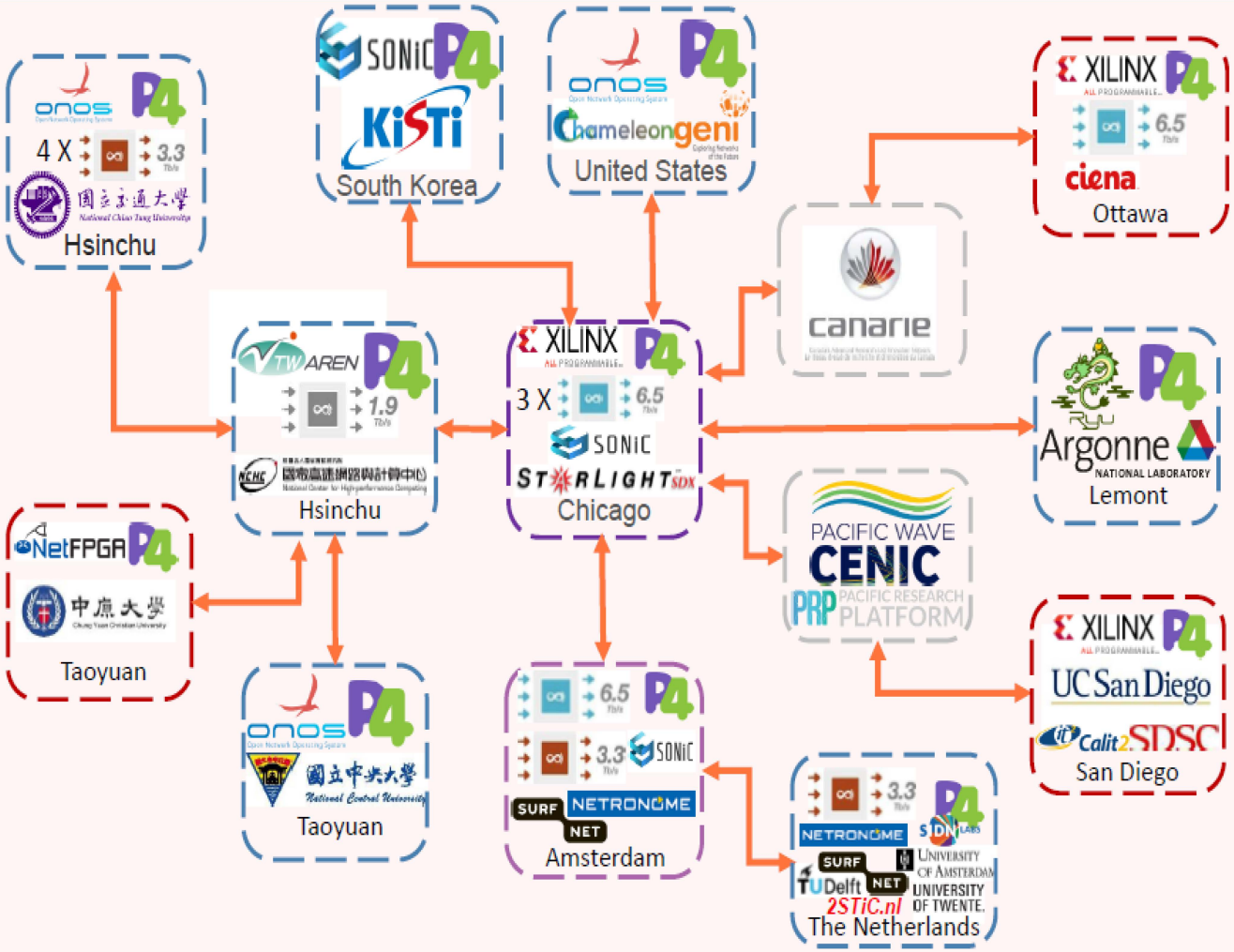
NRE03, NRE04
NRE05, NRE06
NRE10, NRE11
NRE12, NRE13
NRE14
indis104s1



100G Science DTN testbeds Since 2011

NRE08-GRP Service: International P4 Experimental Networks(iP4EN)

SCinet



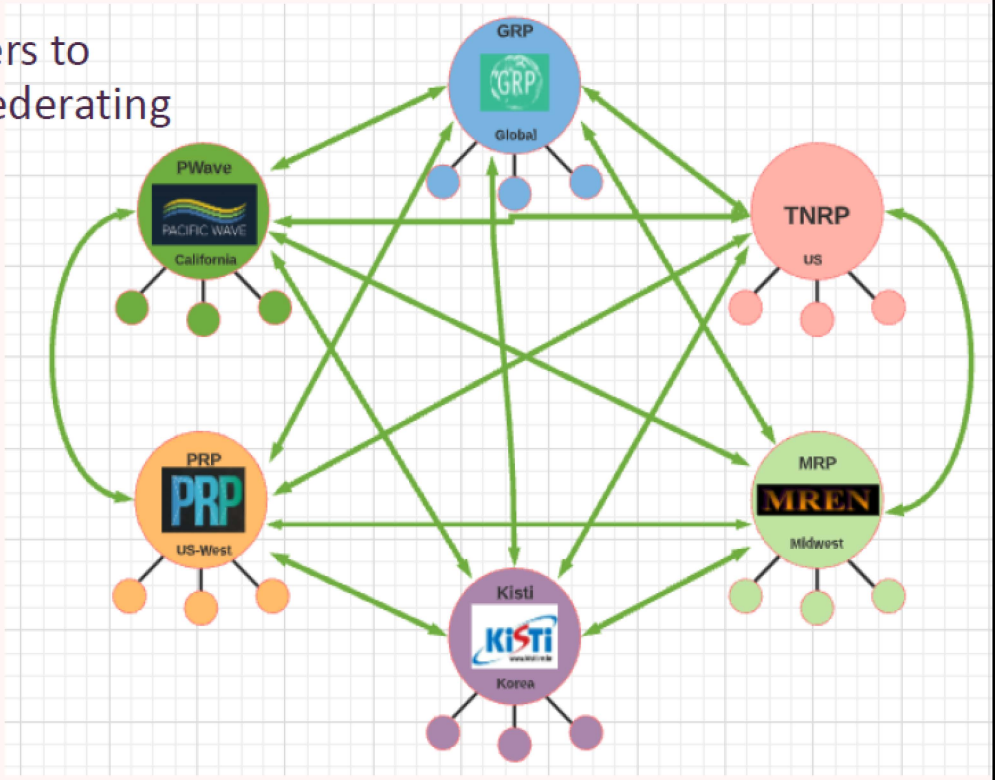
NRE11-GRP Service: Research Platforms Federation Demonstration

Goal: Secure multi-domain resource sharing cross regional, national and international research platforms

Solution: Admiralty. The software enables users to schedule workloads in a different cluster by federating the source and target clusters.

Participants:

- Global Research Platform(GRP)
- Pacific Research Platform(PRP)
- MREN Research platform(MRP)
- Towards National Research Platform(TNRP)
- PacficWave
- KISTI



BigData Express: Toward Predictable, Schedulable, High-performance Data Transfer

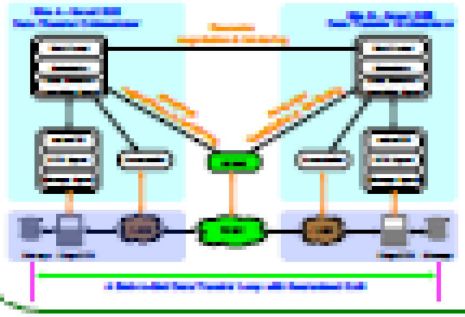
Fermilab

ENERGY

What Is BigData Express?

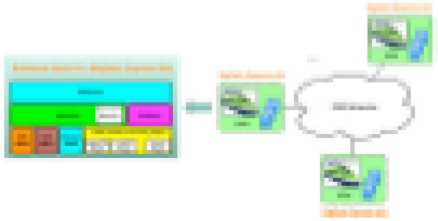
- A schedulable, predictable, and high-performance data transfer service
  - Peer-to-peer, scalable, and extensible data transfer model
  - Visually appealing, easy-to-use web portal
  - High-performance data transfer engine
  - On-demand provisioning of network paths with guaranteed QoS
  - Robust and flexible error handling
  - CILogon-based security
  - A rich set of REST APIs to support scientific workflows
- Project web site: http://bigdataexpress.fnal.gov

End-to-End Transfer Loop

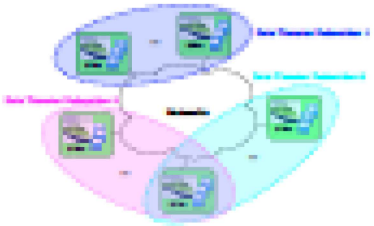


- Application-aware network service
  - On-demand programming
- Fast-provisioning of end-to-end network paths with guaranteed QoS
- Distributed resource negotiation & brokering

Distributed, Peer-to-Peer Model

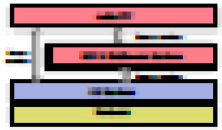


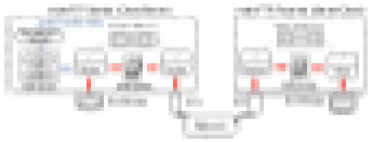
Each site independently provides data transfer service



Flexible data transfer federations

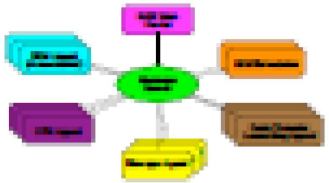
A High-performance Data Transfer Engine – mdtmFTP



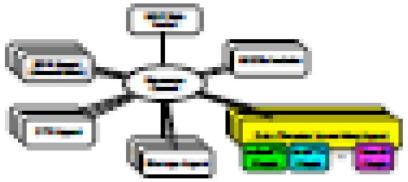


- A pipelined I/O centric design to streamline data transfer
- Multicore-aware data transfer middleware optimizes use of underlying hardware
- Extremely efficient in transferring of Lots Of Small Files

Scalable & Extensible Design



- Scheduler manages site resources through agents
- Use MQTT as message bus



- Extensible Plugin framework to support various data transfer protocols
  - mdtmFTP, GridFTP, XrootD

Miscellaneous

- Apache 2.0 License
- Docker release is available at publicregistry.fnal.gov
- Manuals:
  - Admin: https://bigdataexpress.fnal.gov/admin_manual/index.html
  - Web portal user: https://bigdataexpress.fnal.gov/portal_manual/index.html
- Deployment and collaborations

KISTI KREONET MAX ciena AutoLight KSTAR STARLIGHT NCI

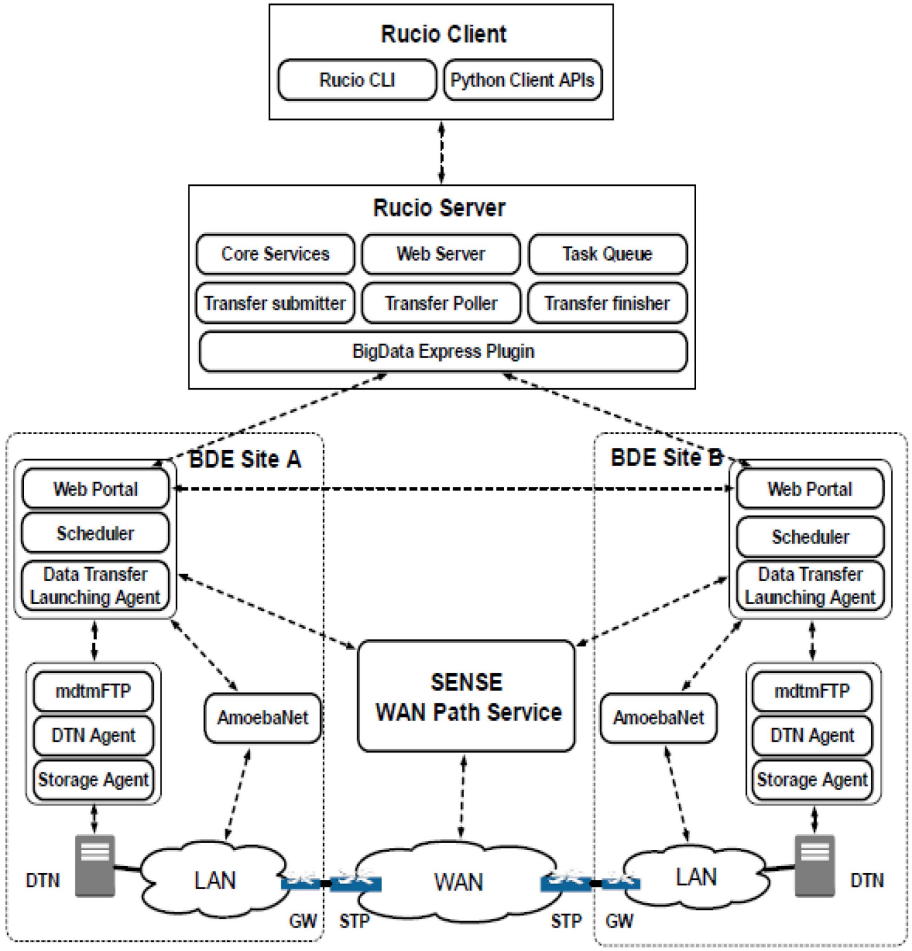
STARLIGHT℠

ROBIN - PI Wenji Wu Fermi National Accelerator Laboratory

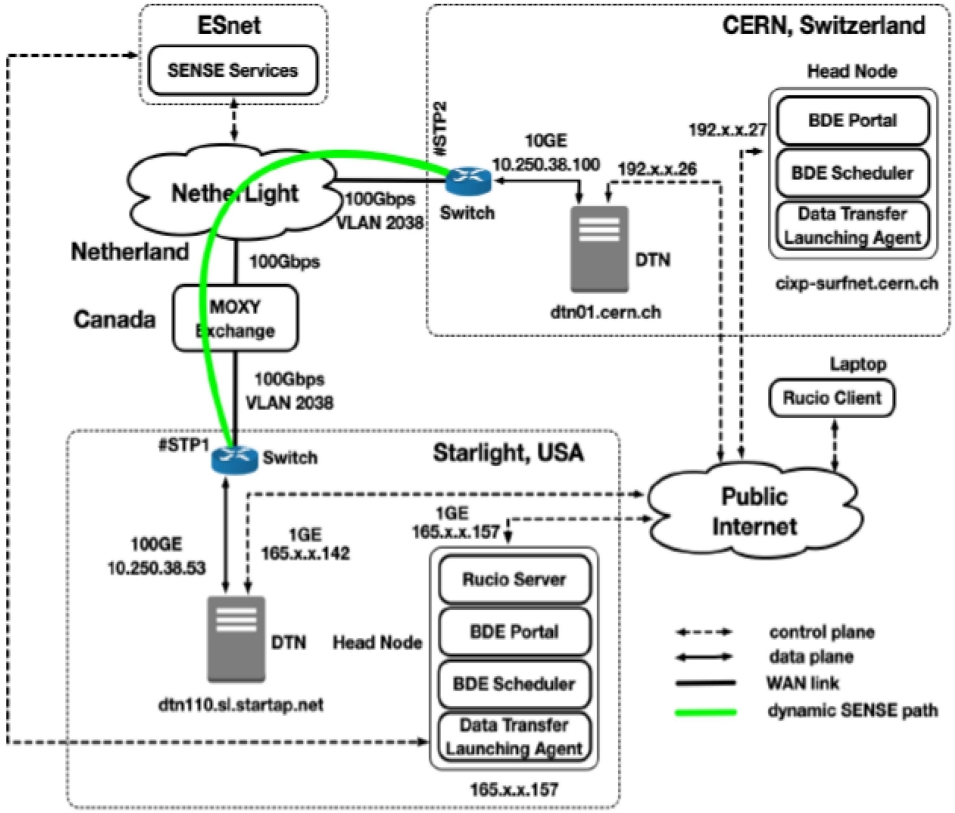
ROBIN (Rucio/BigData Express/SENSE)
A Next Generation High-performance Data Service Platform

Scientific Applications (e.g., LSST, CMS)	
Rucio Data Management Service	- Smart Namespace - Storage Support - Consistency - Easy Integration - AA Support - Proven Track Record
BigData Express High-performance Data Transfer Service	- Peer-to-peer, Scalable, Extensible - End-to-end Performance Optimization - High Performance Data Transfer Engine - CILogon-base Security
SENSE Smart Network Service	- Software Defined-Networking (SDN) for End-to-End Virtual Guaranteed Network - A Highly Intuitive "Intent"-based Interface

ROBIN (Rucio/BigData Express/SENSE)
A Next Generation High-performance Data Service Platform



ROBIN Cross-Atlantic Testbed



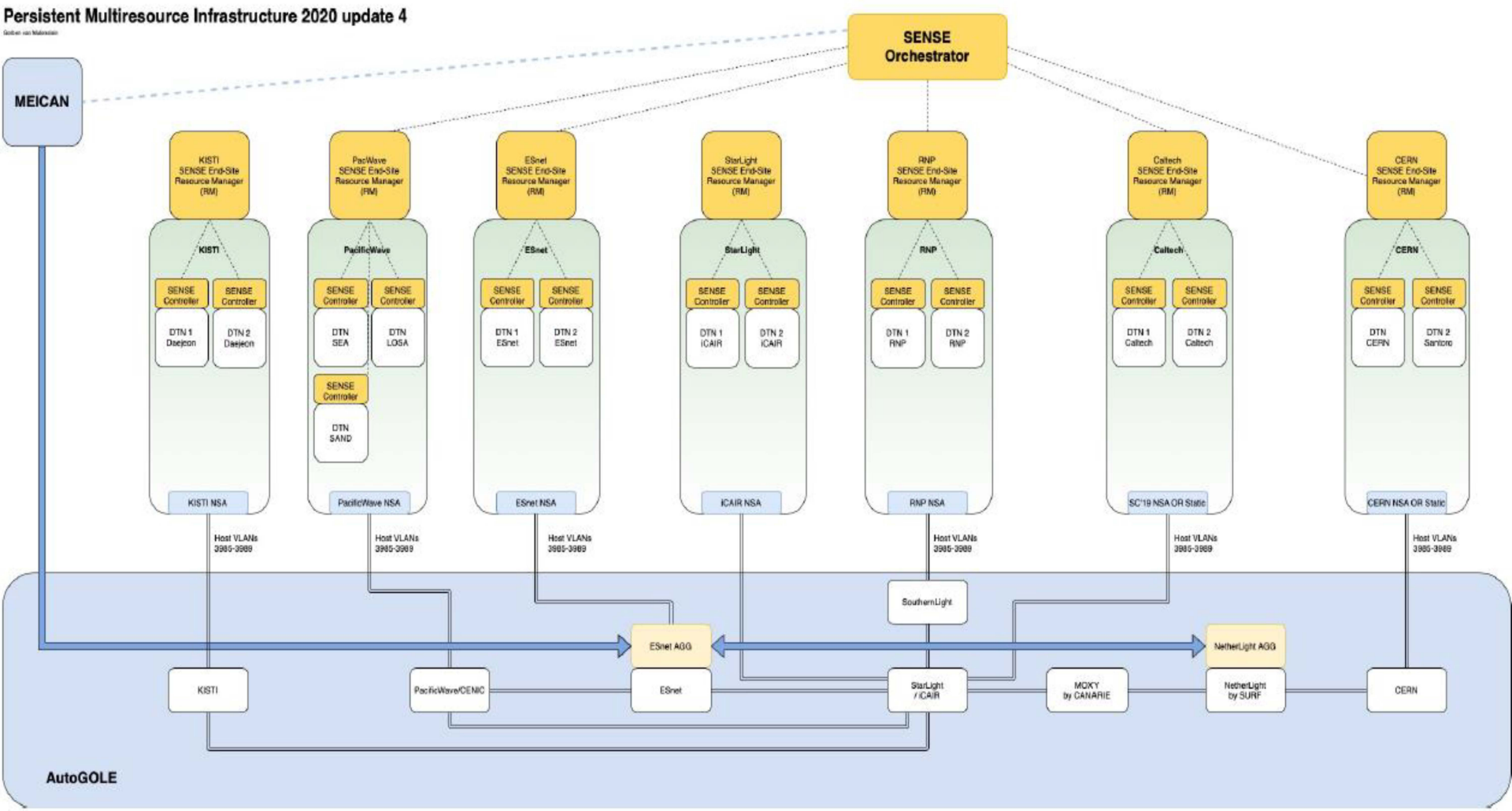
StarLight site:

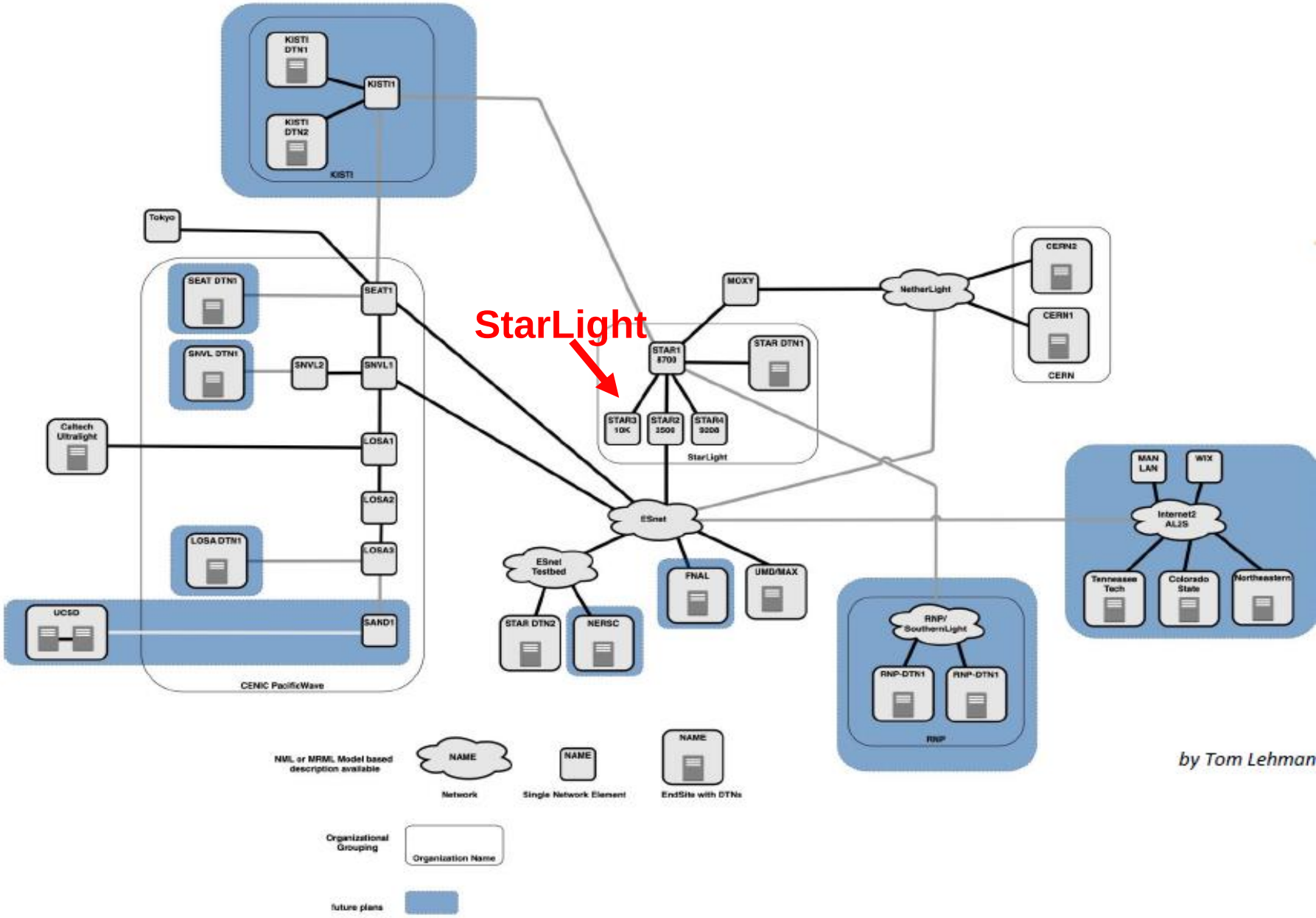
- DTN: *dtn110.sl.startap.net*, with several Intel NVMe drives for data storage, a 100GE Mellanox NIC for data transfer, and a 1G NIC for control.
- Head node: *165.x.x.157,* with a 1G NIC for control.

CERN site:

- DTN: *dtn01.cern.ch*, with a rotational disk for data storage, a 10GE Mellanox NIC for data transfer, and 1G NIC for control.
- Head node: *cixp-urfnet.cern.ch,* with a 1G NIC for control.

The Persistent Multiresource Infrastructure

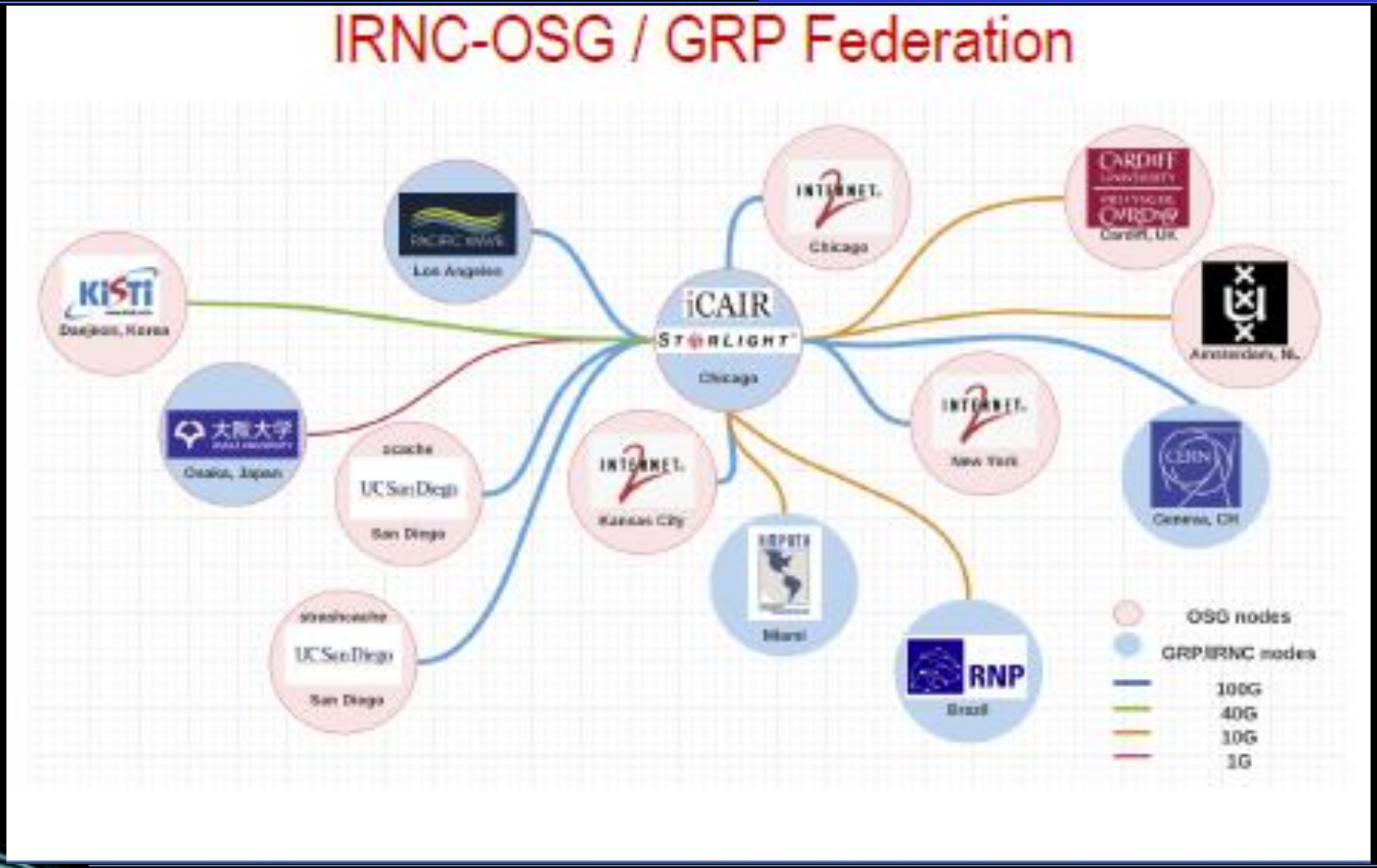




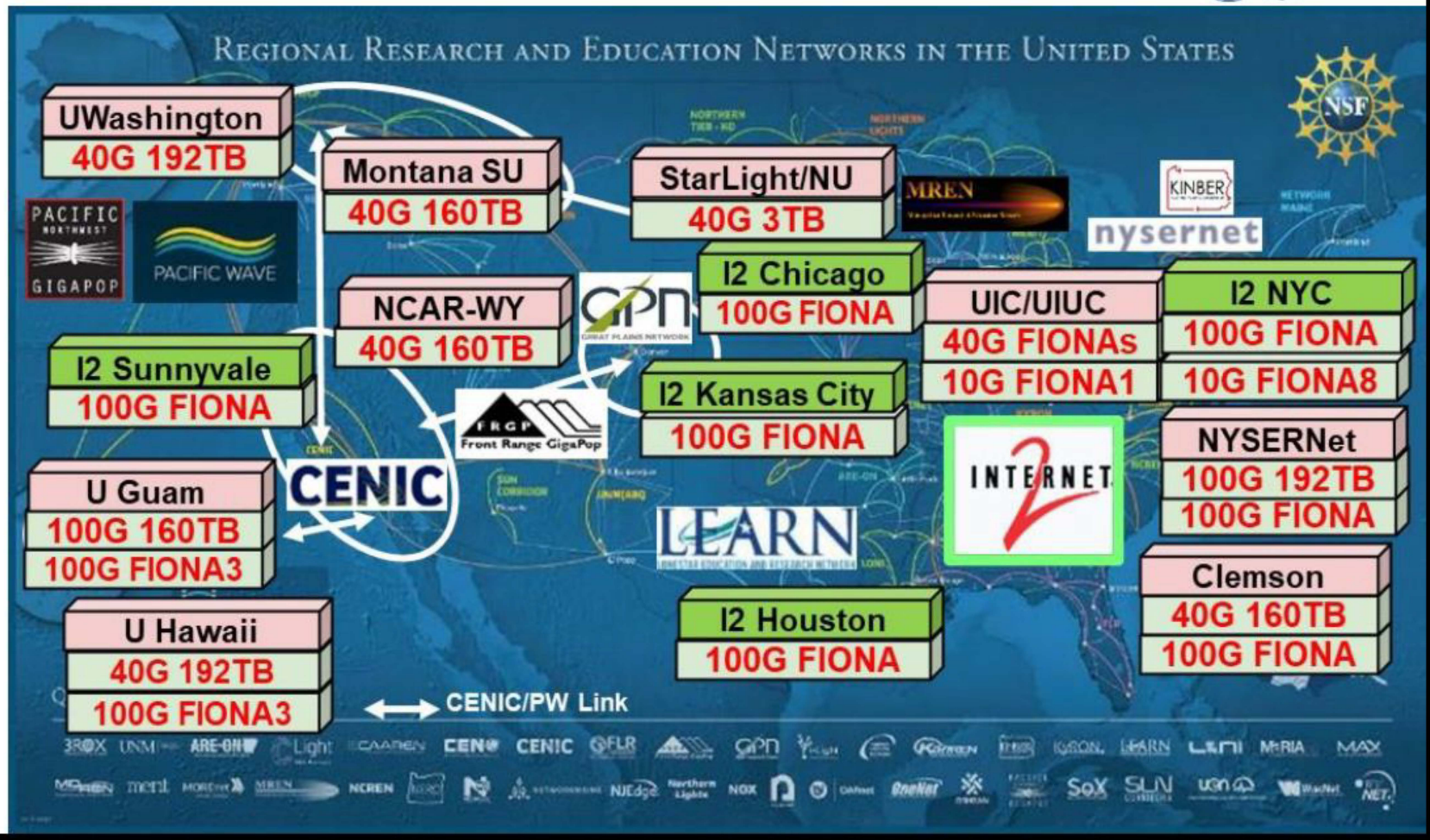
AutoGOLE-SENSE – Tom Lehman

STARLIGHT℠

IRNC-OSG Global Fabric



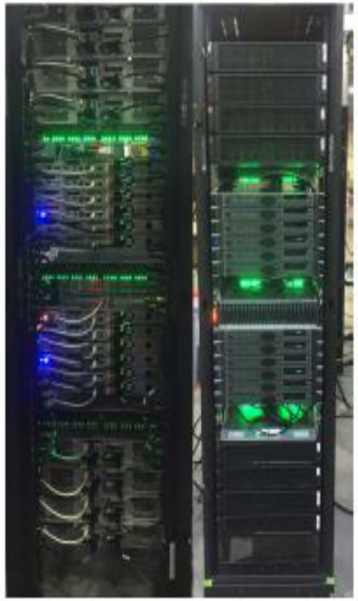
PRP/TNRP's United States (Outside California) Nautilus Hypercluster Connects CENIC and 4 Regionals and 5 Internet2 Sites



Composable Platform as a Service

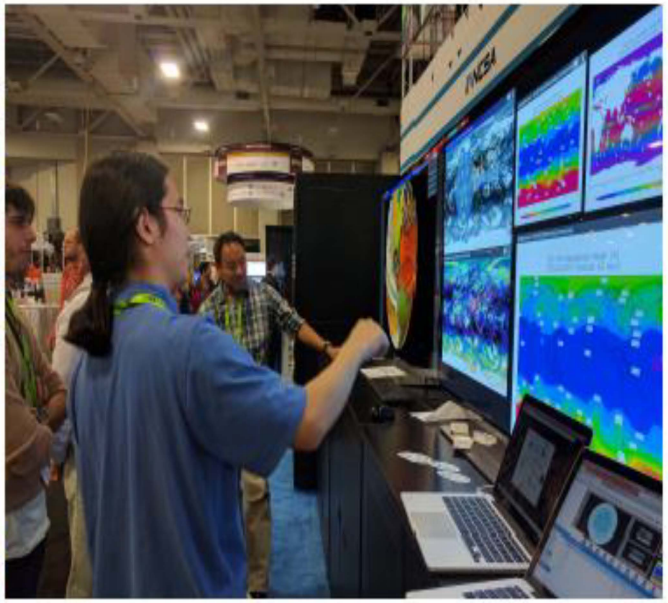
Instrument for Deep Learning & Visualization (COMPaaS DLV)

Using the *Liqid* Composable Infrastructure System for...



...on-demand post-processing, AI, deep-leaning, CFD, and visualization.

...interoperable Science DMZs and Data Transfer Nodes (DTNs) at a national/ international scale, using best strategies from the Pacific Research Platform (PRP) over the GLIF infrastructure.



...SAGE2™, the *de facto* operating system for managing Big Data on tiled-display walls, enabling local and remote collaborators to access, share and interact with a variety of digital media.

evl

Electronic Visualization Laboratory
University of Illinois at Chicago
https://www.evl.uic.edu

NSF award CNS-1828265 to UIC/EVL for COMPaaS DLV
https://www.evl.uic.edu/compaas-dlv

SAGE2 is trademark of the University of Illinois Board of Trustees

Liqid: https://www.liqid.com/
PRP: https://prp.ucsd.edu
GLIF: https://www.glif.is
SAGE2: http://sage2.sagecommons.org/

Chameleon

www.chameleoncloud.org

CHAMELEON: A LARGE SCALE, RECONFIGURABLE EXPERIMENTAL INSTRUMENT FOR COMPUTER SCIENCE

Kate Keahey

Joe Mambretti, Pierre Riteau, Paul Ruth, Dan Stanzione

THE UNIVERSITY OF CHICAGO TACC renci NORTHWESTERN UNIVERSITY NSF

Ilya Baldine PI, RENCI: FABRIC

Next GRP Workshop

- Date: September 13-17, 2021
- Place: Innsbruck Austria, Co Located With eSCience Conference

- Global Research Platform Website:
- www.theglobalresearchplatform.net

THANKS!

- Academia Scinica, AmLight, AARnet, Asia-Pacific Research Platform Consortium, AutoGOLE Consortium, CalTech, CANARIE, CENIC, CERN, Ciena Research Lab, Esnet, Cybera, Fermi National Accelerator Laboratory, Global Research Platform Consortium, International Center for Advanced Internet Research Northwestern University, KISTI, LHC Networking Consortium, NASA Goddard Space Flight Center, Metropolitan Research and Education Network (MREN), MREN Research Platform Consortium (MRP), Mid-Atlantic Crossroads (MAX), Naval Research Laboratory, National Energy Research Scientific Computing Center (NERSC), National Science Foundation, Pacific Research Platform, Pacific Wave, RNP, SingAREN, StarLight International/National Communications Exchange Facility Consortium, SURFnet, TWAREN, University of California San Diego, University of Illinois Chicago, University of Amsterdam, et al

www.startap.net/starlight

Thanks to the NSF, DOE, NASA, NIH, DARPA
Universities, National Labs, International & Industrial Partners,
and Other Supporters

STARLIGHT℠

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov