DTN-as-a-Service At Starlight

A GRP Prototype Service

Presenter: Jim Chen

Se-young Yu, Xiao Wang, Fei Yeh, Joe Mambretti
International Center for Advanced Internet Research
Northwestern University
Starlight

Sep 18 2019, 1st GRP workshop, San Diego CA

iCAIR

STARLIGHT SDX

Overview

1. Overview
2. Global Research Platform(GRP) prototype services:
    - GRP cluster with Kubernetes
    - DTN-as-a-Service for GRP
    - International P4 Experiment Networks
3. DTN-as-a-Service at Starlight overview
4. DTN options at Starlight and GRP partner sites
5. Starlight DTN-as-a-Service software stack
6. Summary, Q&A

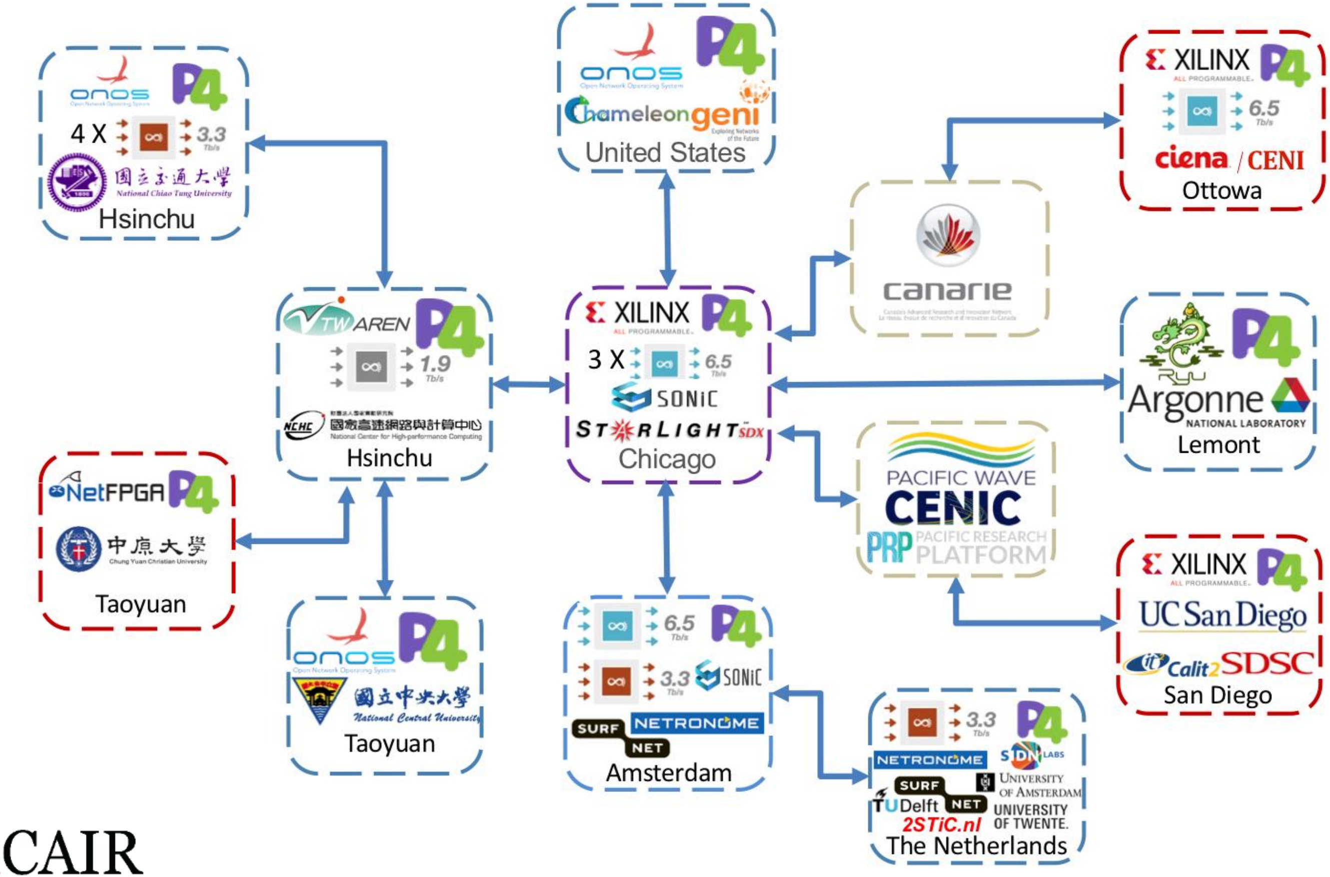
GRP Prototype Services from Starlight

September 2019

- DTN-as-a-Service in Starlight and partner sites
- International P4 Experiment Networks
- Global Research Platform Cluster Environment

- Software stack distribution to support GRP prototypes

International P4 Experimental Networks (i-P4EN)



EUROP4 workshop:
Sep 23 2019,
Cambridge U.K.

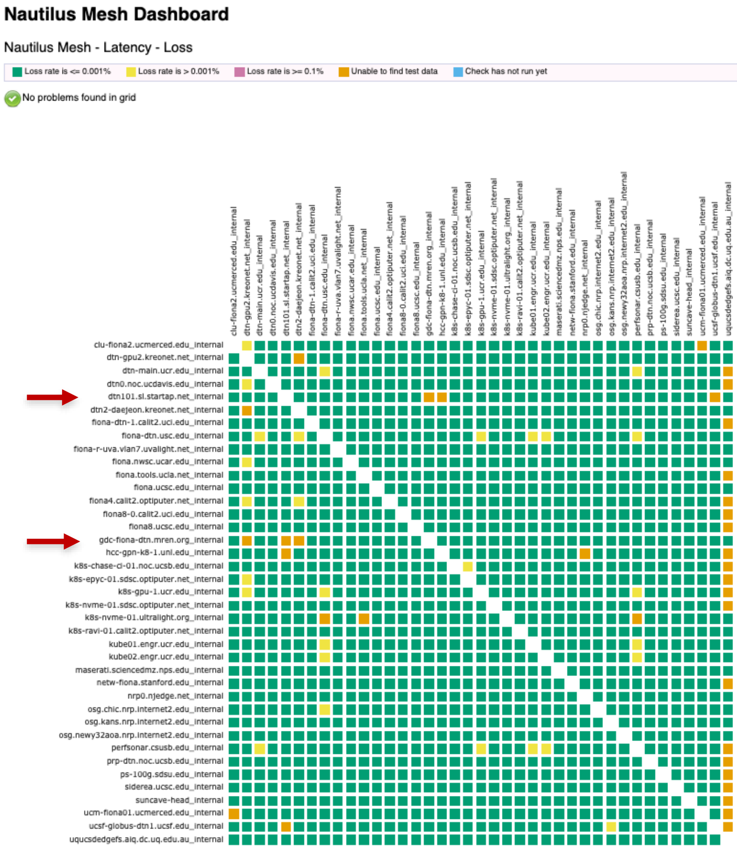
(1) P4MT: Multi-Tenant Support Prototype for International P4 Testbed.

(2) Sketch-based Entropy Estimation for Network Traffic Analysis using Programmable Data Plane.

iCAIR

STARLIGHT SDX

PRP/TNRP,MREN and AutoGOLE Research Platform



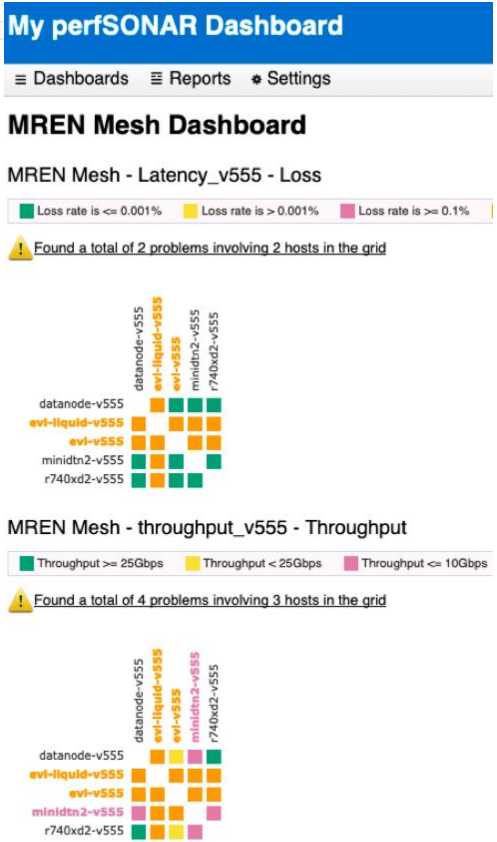
Nautilus Mesh Dashboard

Nautilus Mesh - Latency - Loss

🟩 Loss rate is <= 0.001% 🟨 Loss rate is > 0.001% 🟪 Loss rate is >= 0.1% 🟧 Unable to find test data 🟦 Check has not run yet

✓ No problems found in grid

	clu-fiona2.ucmerced.edu_internal	dtn-gpu2.kreonet.net_internal	dtn-main.ucr.edu_internal	dtn0.noc.ucdavis.edu_internal	dtn101.sl.startap.net_internal	dtn2-daejeon.kreonet.net_internal	fiona-dtn-1.calit2.uci.edu_internal	fiona-dtn.usc.edu_internal	fiona-r-uva.vlan7.uvalight.net_internal	fiona.nwsc.ucar.edu_internal	fiona.tools.ucla.net_internal	fiona.ucsc.edu_internal	fiona4.calit2.optiputer.net_internal	fiona8-0.calit2.uci.edu_internal	fiona8.ucsc.edu_internal	gdc-fiona-dtn.mren.org_internal	hcc-gpn-k8-1.unl.edu_internal	k8s-chase-ci-01.noc.ucsb.edu_internal	k8s-epyc-01.sdsc.optiputer.net_internal	k8s-gpu-1.ucr.edu_internal	k8s-nvme-01.sdsc.optiputer.net_internal	k8s-nvme-01.ultralight.org_internal	k8s-ravi-01.calit2.optiputer.net_internal	kube01.engr.ucr.edu_internal	kube02.engr.ucr.edu_internal	maserati.sciencedmz.nps.edu_internal	netw-fiona.stanford.edu_internal	nrp0.njedge.net_internal	osg.chic.nrp.internet2.edu_internal	osg.kans.nrp.internet2.edu_internal	osg.newy32aoa.nrp.internet2.edu_internal	perfsonar.csusb.edu_internal	prp-dtn.noc.ucsb.edu_internal	ps-100g.sdsu.edu_internal	siderea.ucsc.edu_internal	suncave-head_internal	ucm-fiona01.ucmerced.edu_internal	ucsf-globus-dtn1.ucsf.edu_internal	uqucsdedgefs.aiq.dc.uq.edu.au_internal
clu-fiona2.ucmerced.edu_internal		🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟩	🟩
dtn-gpu2.kreonet.net_internal	🟩		🟩	🟩	🟩	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
dtn-main.ucr.edu_internal	🟩	🟩		🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟧
dtn0.noc.ucdavis.edu_internal	🟩	🟨	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
dtn101.sl.startap.net_internal	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟩
dtn2-daejeon.kreonet.net_internal	🟩	🟧	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
fiona-dtn-1.calit2.uci.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
fiona-dtn.usc.edu_internal	🟩	🟩	🟨	🟩	🟩	🟨	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟨	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩
fiona-r-uva.vlan7.uvalight.net_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
fiona.nwsc.ucar.edu_internal	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
fiona.tools.ucla.net_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
fiona.ucsc.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
fiona4.calit2.optiputer.net_internal	🟩	🟨	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
fiona8-0.calit2.uci.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
fiona8.ucsc.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
gdc-fiona-dtn.mren.org_internal	🟩	🟧	🟩	🟩	🟧	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
hcc-gpn-k8-1.unl.edu_internal	🟩	🟩	🟩	🟩	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
k8s-chase-ci-01.noc.ucsb.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
k8s-epyc-01.sdsc.optiputer.net_internal	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
k8s-gpu-1.ucr.edu_internal	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟧
k8s-nvme-01.sdsc.optiputer.net_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
k8s-nvme-01.ultralight.org_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟩	🟩	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩
k8s-ravi-01.calit2.optiputer.net_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
kube01.engr.ucr.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩
kube02.engr.ucr.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩
maserati.sciencedmz.nps.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
netw-fiona.stanford.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧
nrp0.njedge.net_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
osg.chic.nrp.internet2.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
osg.kans.nrp.internet2.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
osg.newy32aoa.nrp.internet2.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
perfsonar.csusb.edu_internal	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟨	🟨	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟧
prp-dtn.noc.ucsb.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟧
ps-100g.sdsu.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟧
siderea.ucsc.edu_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟧
suncave-head_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟧
ucm-fiona01.ucmerced.edu_internal	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟧
ucsf-globus-dtn1.ucsf.edu_internal	🟩	🟩	🟩	🟩	🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟧
uqucsdedgefs.aiq.dc.uq.edu.au_internal	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	



My perfSONAR Dashboard

MREN Mesh Dashboard

MREN Mesh - Latency_v555 - Loss

🟩 Loss rate is <= 0.001% 🟨 Loss rate is > 0.001% 🟪 Loss rate is >= 0.1%

⚠ Found a total of 2 problems involving 2 hosts in the grid

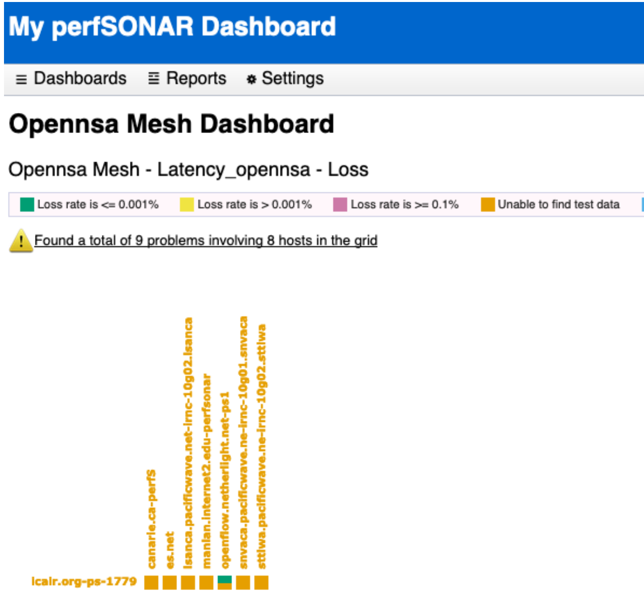
	datanode-v555	evl-liquid-v555	evl-v555	minidtn2-v555	r740xd2-v555
datanode-v555		🟧	🟩	🟩	🟩
evl-liquid-v555	🟧		🟧	🟧	🟧
evl-v555	🟧	🟧		🟧	🟧
minidtn2-v555	🟩	🟧	🟩		🟩
r740xd2-v555	🟩	🟧	🟩	🟩	

MREN Mesh - throughput_v555 - Throughput

🟩 Throughput >= 25Gbps 🟨 Throughput < 25Gbps 🟪 Throughput <= 10Gbps

⚠ Found a total of 4 problems involving 3 hosts in the grid

	datanode-v555	evl-liquid-v555	evl-v555	minidtn2-v555	r740xd2-v555
datanode-v555		🟧	🟨	🟪	🟩
evl-liquid-v555	🟧		🟧	🟧	🟧
evl-v555	🟧	🟧		🟧	🟧
minidtn2-v555	🟪	🟧	🟧		🟪
r740xd2-v555	🟩	🟧	🟨	🟪	



My perfSONAR Dashboard

Opennsa Mesh Dashboard

Opennsa Mesh - Latency_opennsa - Loss

🟩 Loss rate is <= 0.001% 🟨 Loss rate is > 0.001% 🟪 Loss rate is >= 0.1% 🟧 Unable to find test data

⚠ Found a total of 9 problems involving 8 hosts in the grid

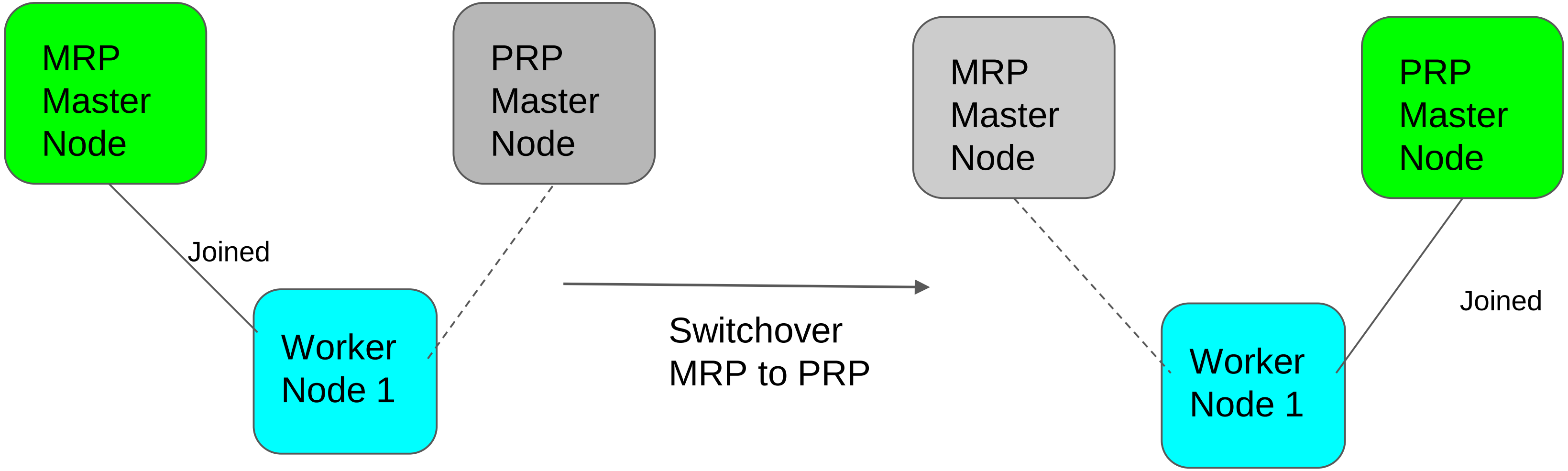
	canarie.ca-perfS	es.net	lsanca.pacificwave.net-lrnc-10g02.lsanca	manlan.internet2.edu-perfsonar	openflow.netherlight.net-ps1	snvaca.pacificwave.ne-lrnc-10g01.snvaca	sttlwa.pacificwave.ne-lrnc-10g02.sttlwa
icair.org-ps-1779	🟧	🟧	🟧	🟧	🟩🟧	🟧	🟧

iCAIR

STARLIGHT SDX

Advanced feature: Multi-Cluster Controller

The Multi-cluster controller is developed for worker nodes. The goal is to enable worker nodes to dynamically participating a cluster on-demand. This is one of virtualization solutions for worker nodes to participating multi-clusters.



Starlight DTN-as-a-Service Highlights

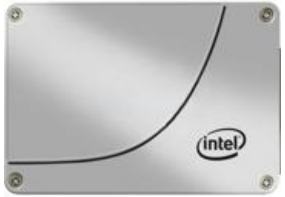
Starlight DTNaaS platform provides:

1. NUMA-aware task and process autonomous configuration
2. Autonomous optimization for the underlying hardware and software system
3. Modular data transfer system integration platform
4. Support data access with NVMe over Fabrics
5. Science workflow user interface for network provisioning with NSI/OpenNSA
6. A monitoring system for high-performance data transfer

Starlight DTN-as-a-Service Benefits

- Enabling users to move data without any knowledge for the underlying infrastructure.
- A platform for autonomous configuration and optimization for the data transfer using DTNs.
- Support operation in Docker, Singularity and K8s with Docker.
- Support NVMe over Fabrics for access remote storage as a local device.
- Users can evaluate the data movement in real-time, reconfigure the system, and change the transfer tools as required.
- Modular design, implemented on Jupyter + Python, perfect for science research workflow integration.

SC16: Supermicro 24X NVMe SuperServer



Option A: Intel P3700 800G X 16

or soon to be Intel P4600

Option B: SamSung 950 Pro 512G/960 Pro 1T or 2T+ M.2 to U.2 Adopter X16

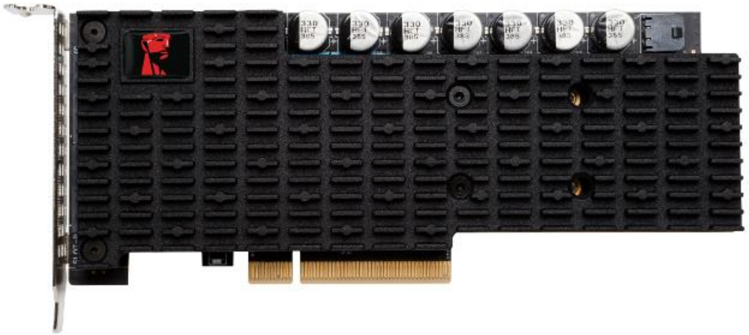
iCAIR

STARLIGHT SDX

SC17: Scinet DTN EchoStream 1U

2 X Mellanox ConnectX-4 100GE

4 X Liquid/Kingston NVMe PCI-e X8 AIC



iCAIR

STARLIGHT SDX

SC17: SDX Scalable DTN+AI Prototype Solution

NVMe A: Intel P3700 800G X 8

NVMe B: Samsung 960 Pro 1T X 8

+ M.2 to U.2 Adopter

GPU: NVIDIA P1000 X 2 + V100 X 1

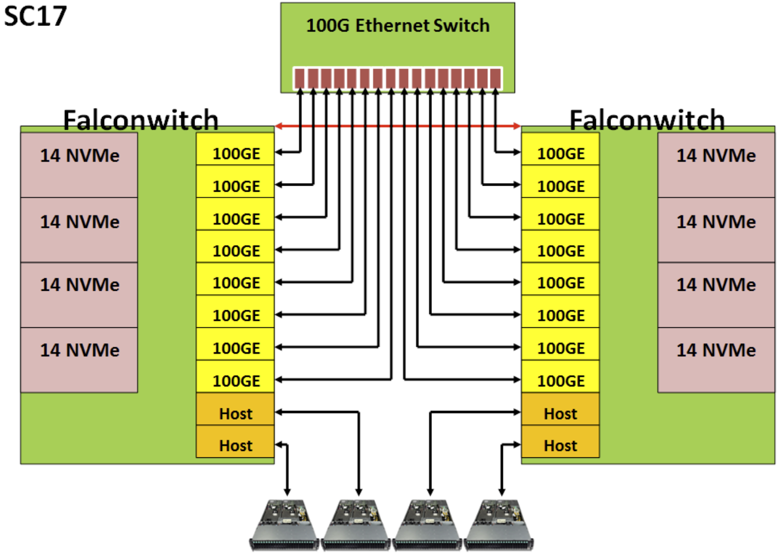
Host node: SuperWorkstation 7048GR-TR

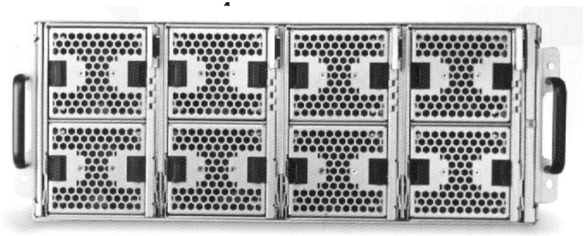
2 X Mellanox ConnectX-5

2 X Intel E5-2667 V4

100GE cards

192G RAM

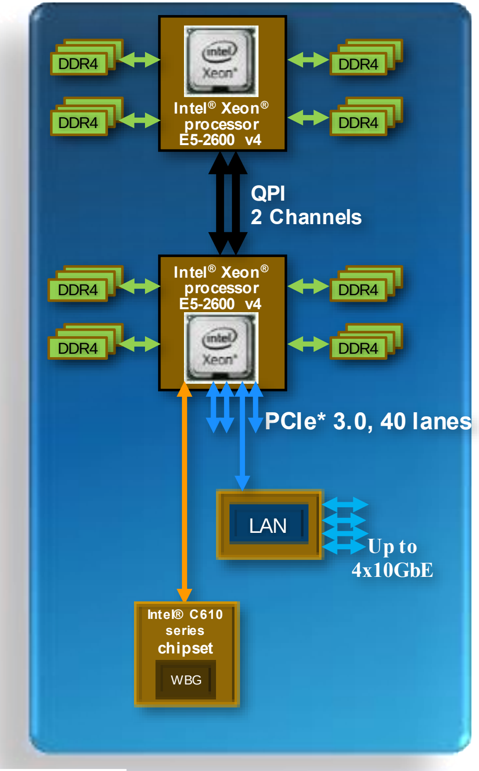




iCAIR

STARLIGHT SDX

SC18 Scinet DTN: Dell 14G R740_XD Solution



2 X Intel Xeon Gold 3.0+ GHz CPUs

2 X Mellanox ConnectX-4 100GE

4 X Liquid/Kingston 1.6T/3.2T NVMe PCI-e X8 AIC

iCAIR

STARLIGHT SDX

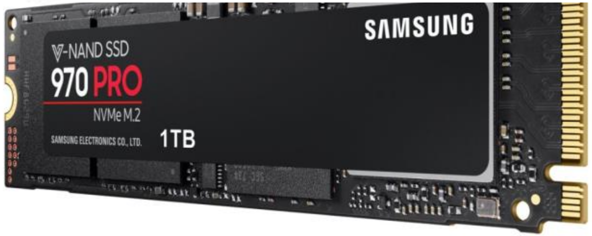
SC19: Scinet DTN AMD Supermicro 3U

2 X Mellanox ConnectX-5 100GE

4 X Quattro 400 M.2 NVMe Adapter

16 X Samsung NVMe M.2 970 Pro 1T

AMD EPYC 7371 16C 3.1/3.6GHz



iCAIR

STARLIGHT SDX

SEAIP Data Movers 1G/10G DTN

SEAIP

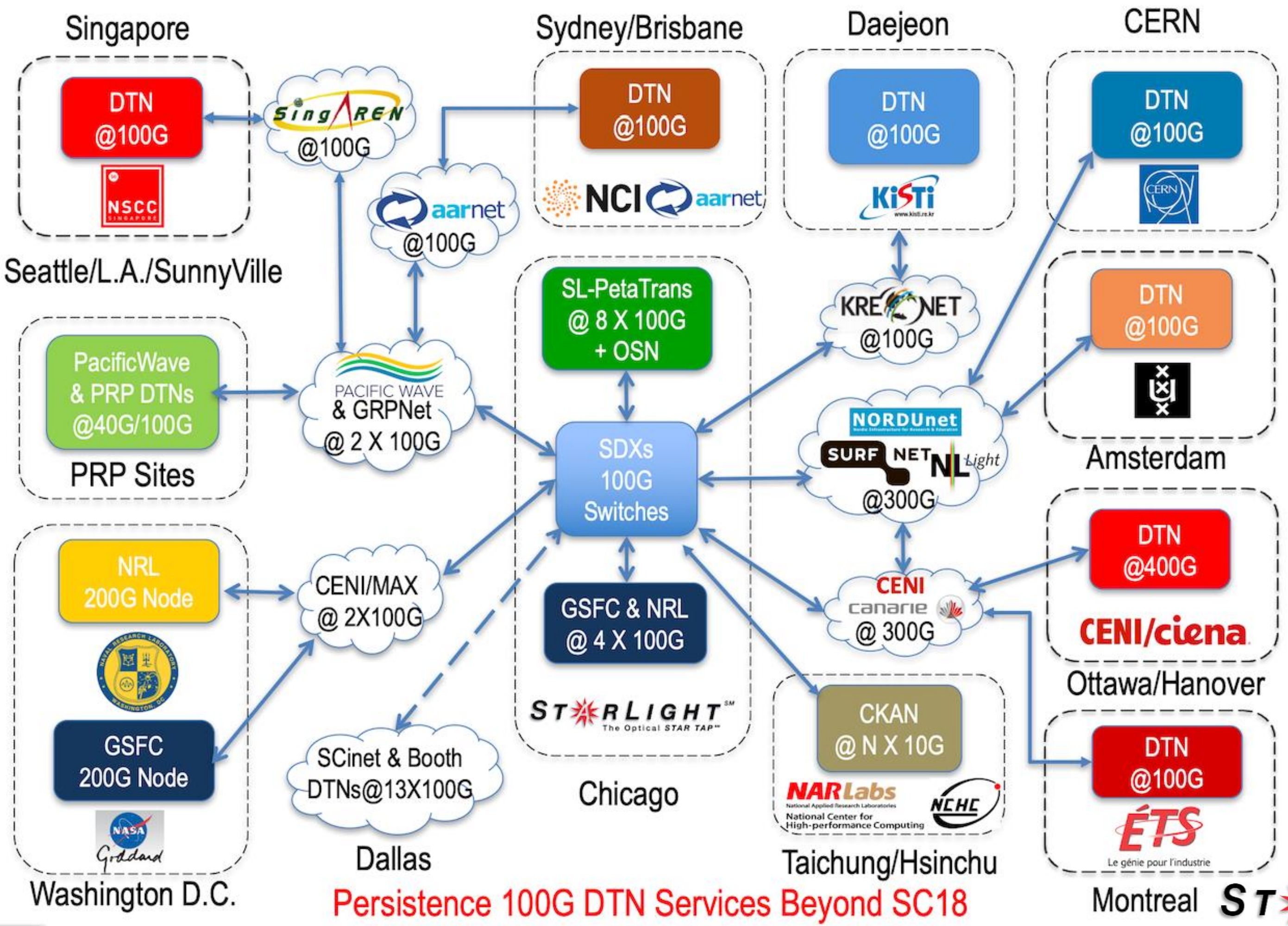
Intel NUC8i5BEK

i5-8259U 3.8GHz quad-core CPU

1 X Samsung NVMe M.2 970 Pro 1T

Thunderbolt 3 - 10GE Converter

PetaTrans: Petascale Sciences Data Transfer

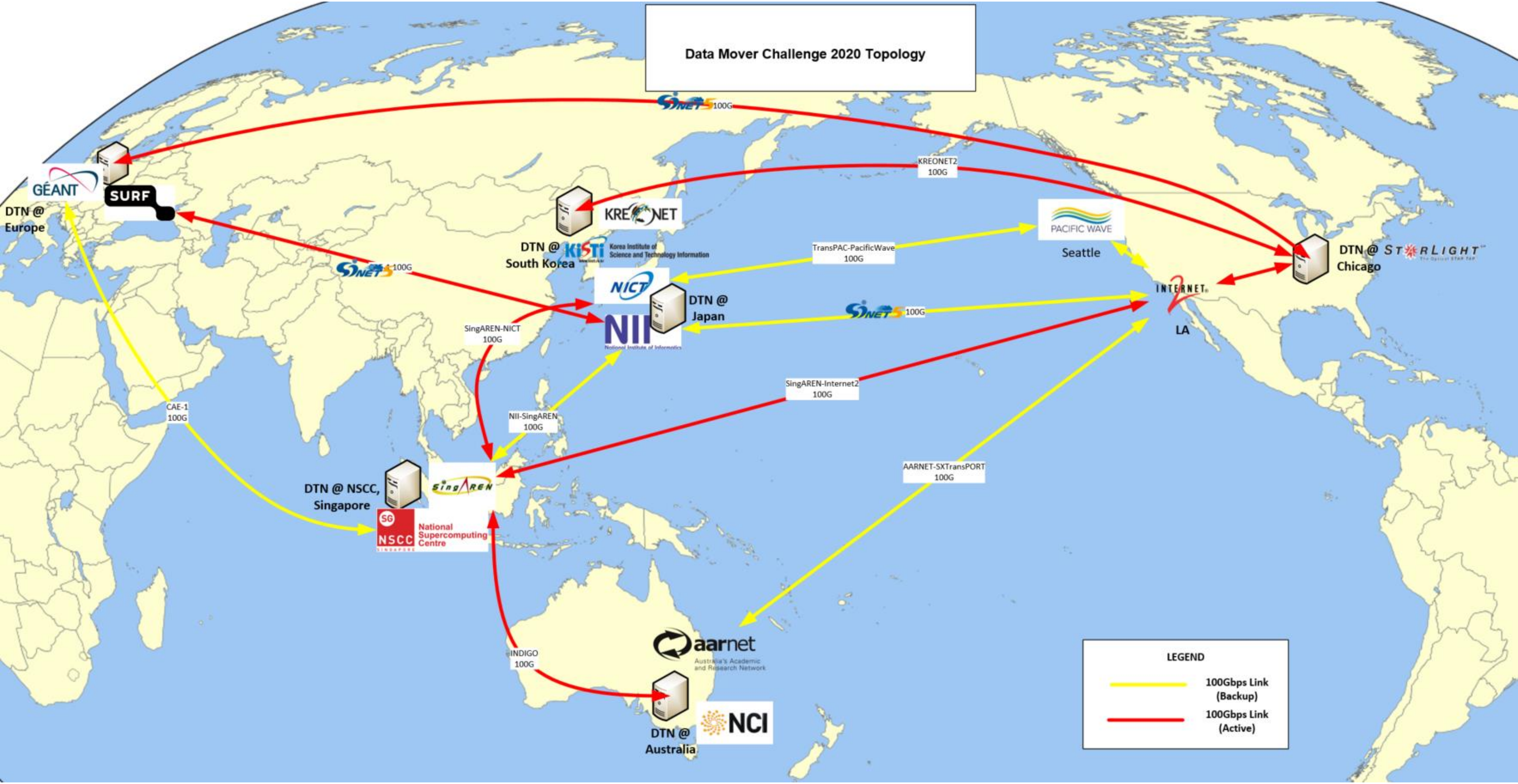


iCAIR

Persistence 100G DTN Services Beyond SC18

STARLIGHT SDX

Data Mover Challenge 2020 Topology



SCA19: Starlight DTNaaS Software Stack

- Optimize the transfer performance based on the machine configuration
- Provide functions to automate data transfer
- Set up and tear down transfer-tool environment supported on the DTNs
- Modular component to support additional data transfer tools and additional science workflow
- For SCA19 DMC, nuttcp transfer tool is used for disk-to-disk, Built-in iperf3 is used for memory-to-memory transfer
- Work-flow controller implemented in Jupyter to enable easy to integrate research & collaboration

SCAsia 2019
Supercomputing

Gathering the Best of HPC in Asia

SG
NSCC
SINGAPORE

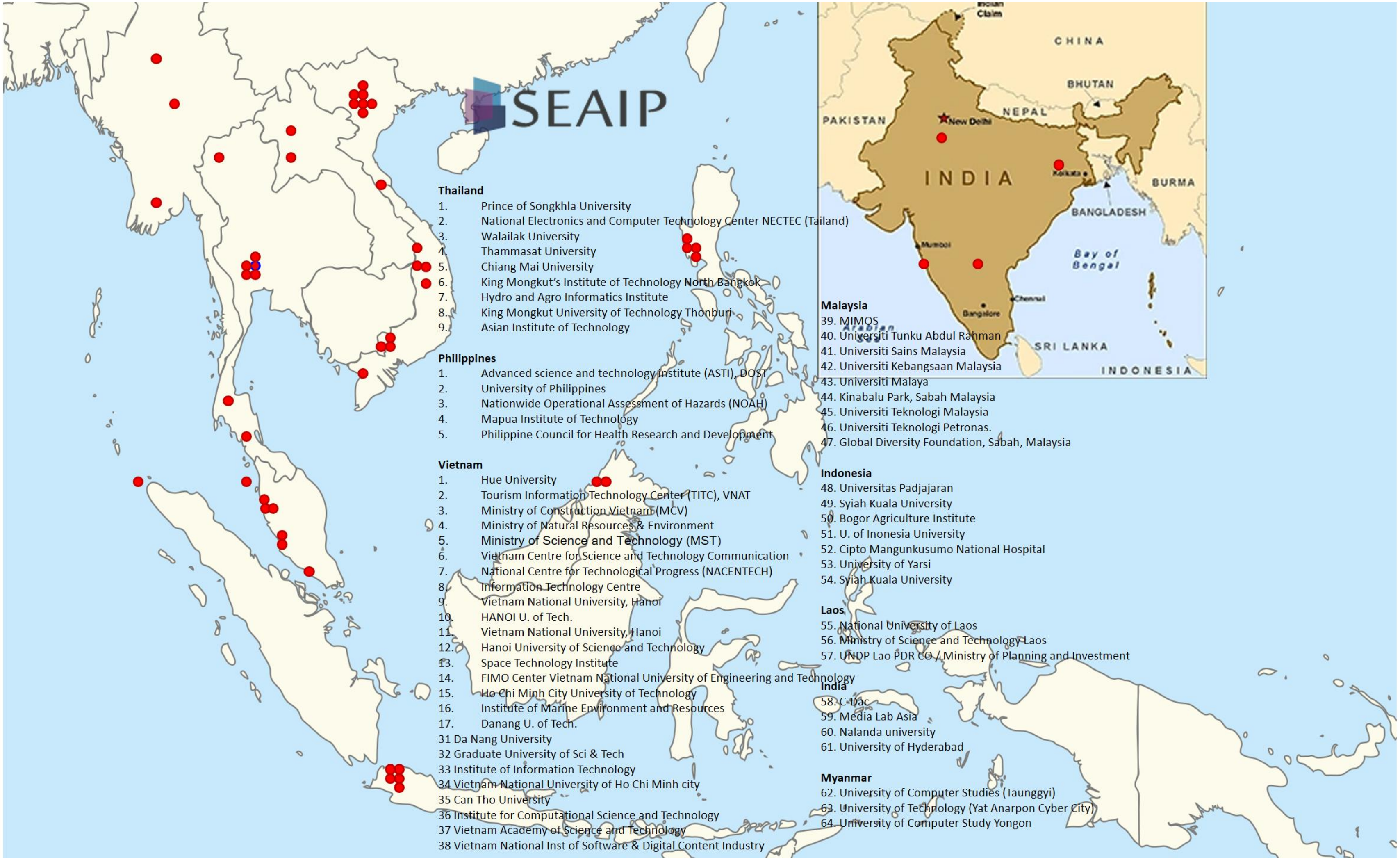
SupercomputingAsia 2019
Data Mover Challenge
Most Innovative Solution

StarLight / iCAIR

SCA19 & 20: SEAIP Data Mover Team

- SEAIP: Southeast Asia International joint-research and training Program, Established 10 years.
- SEAIP Team Project Objectives: Initiate Data Mover Service Collaboration and Enable DTN Services At Different Sites/Countries.
- Project Team Established During SEAIP2018 Workshop, Nov 26-30, 2018
- Team Lead: Steven Shiau(NCHC), Co-Leads: Jim Chen (iCAIR), Te-Lung Liu(NCHC) With 15+ Participants From 6 Countries.
- Proposed Innovations: Gateway For Different Speed DTNs, CloneZilla Data Transfer Service for Bare-Metal Data Mover.

SEAIP



Thailand

1. Prince of Songkhla University
2. National Electronics and Computer Technology Center NECTEC (Tailand)
3. Walailak University
4. Thammasat University
5. Chiang Mai University
6. King Mongkut's Institute of Technology North Bangkok
7. Hydro and Agro Informatics Institute
8. King Mongkut University of Technology Thonburi
9. Asian Institute of Technology

Philippines

1. Advanced science and technology Institute (ASTI), DOST
2. University of Philippines
3. Nationwide Operational Assessment of Hazards (NOAH)
4. Mapua Institute of Technology
5. Philippine Council for Health Research and Development

Vietnam

1. Hue University
2. Tourism Information Technology Center (TITC), VNAT
3. Ministry of Construction Vietnam (MCV)
4. Ministry of Natural Resources & Environment
5. Ministry of Science and Technology (MST)
6. Vietnam Centre for Science and Technology Communication
7. National Centre for Technological Progress (NACENTECH)
8. Information Technology Centre
9. Vietnam National University, Hanoi
10. HANOI U. of Tech.
11. Vietnam National University, Hanoi
12. Hanoi University of Science and Technology
13. Space Technology Institute
14. FIMO Center Vietnam National University of Engineering and Technology
15. Ho Chi Minh City University of Technology
16. Institute of Marine Environment and Resources
17. Danang U. of Tech.

31 Da Nang University
32 Graduate University of Sci & Tech
33 Institute of Information Technology
34 Vietnam National University of Ho Chi Minh city
35 Can Tho University
36 Institute for Computational Science and Technology
37 Vietnam Academy of Science and Technology
38 Vietnam National Inst of Software & Digital Content Industry

Malaysia

39. MIMOS
40. Universiti Tunku Abdul Rahman
41. Universiti Sains Malaysia
42. Universiti Kebangsaan Malaysia
43. Universiti Malaya
44. Kinabalu Park, Sabah Malaysia
45. Universiti Teknologi Malaysia
46. Universiti Teknologi Petronas.
47. Global Diversity Foundation, Sabah, Malaysia

Indonesia

48. Universitas Padjajaran
49. Syiah Kuala University
50. Bogor Agriculture Institute
51. U. of Inonesia University
52. Cipto Mangunkusumo National Hospital
53. University of Yarsi
54. Syiah Kuala University

Laos

55. National University of Laos
56. Ministry of Science and Technology Laos
57. UNDP Lao PDR CO / Ministry of Planning and Investment

India

58. C-DAC
59. Media Lab Asia
60. Nalanda university
61. University of Hyderabad

Myanmar

62. University of Computer Studies (Taunggyi)
63. University of Technology (Yat Anarpon Cyber City)
64. University of Computer Study Yongon

X-NET: SCinet Data Transfer Node(DTN) Service

TEAM MEMBERS

Jim Chen	NWU/STARLight
Gonzalo Rodrigo	Apple/LBL
Ana Giannakou	LBL
Eric Pouyoul	ESnet
Fei Yeh	NWU/STARLight
Se-Young Yu	NWU/STARLight

TO DO:

1) Develop 100G network fiber/link/vlan/route verification procedures with a portable tester to shorten set up time and improve readiness.
2) Prototype user experiment environment isolation & management solutions: Docker/ Kubernetes/Rancher/VM, also plan to evaluate other Docker Integration
3) Design AI-Enabled DTN use case and workflow prototype

Related & Supported Paper:

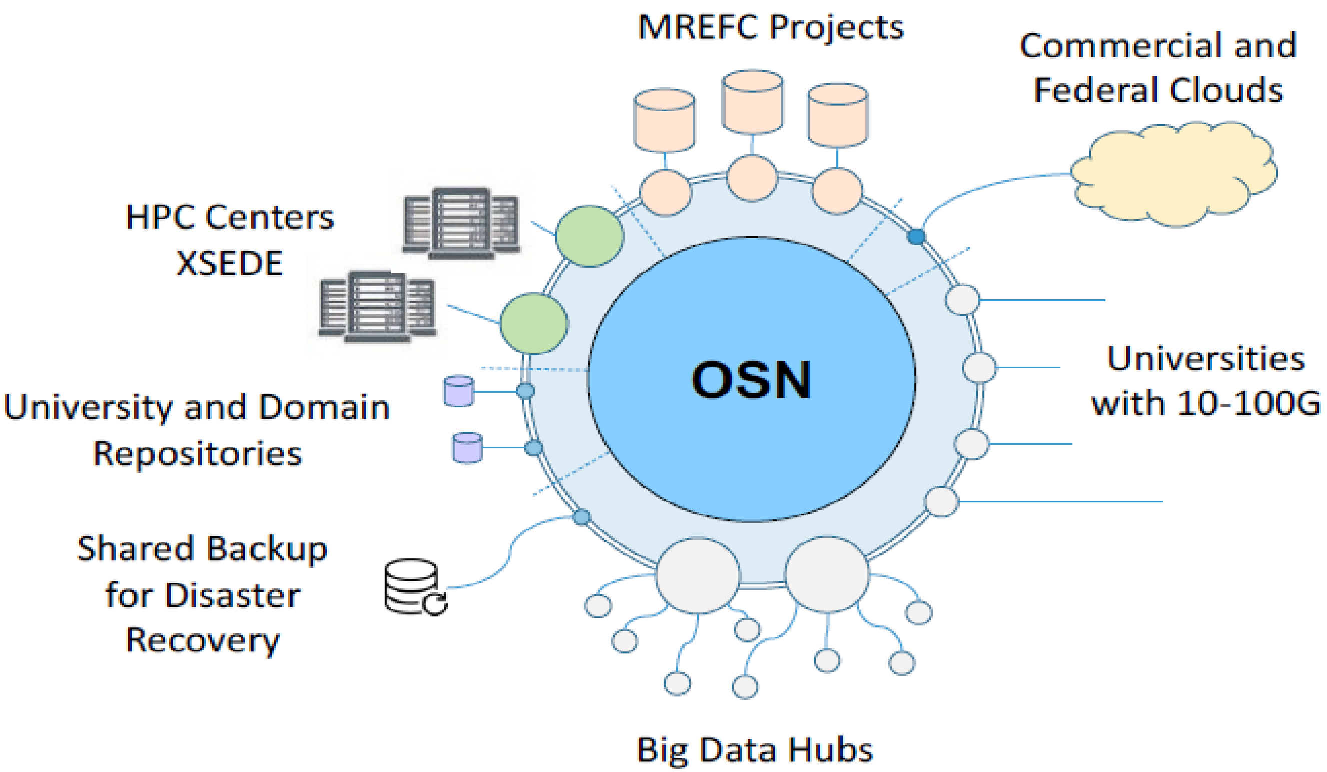
1) "Analysis of CPU Pinning and Storage Configuration in 100 Gbps Network Data Transfer"
   -Se-Young Yu & others.
2) "BigData Express: Toward Schedulable, Predictable, and High-performance Data Transfer"
   -Wenji Wu & other
3) "Flowzilla: A methodology for Detecting Data Transfer Anomalies in Research Networks."
   -Anna Giannakou & others

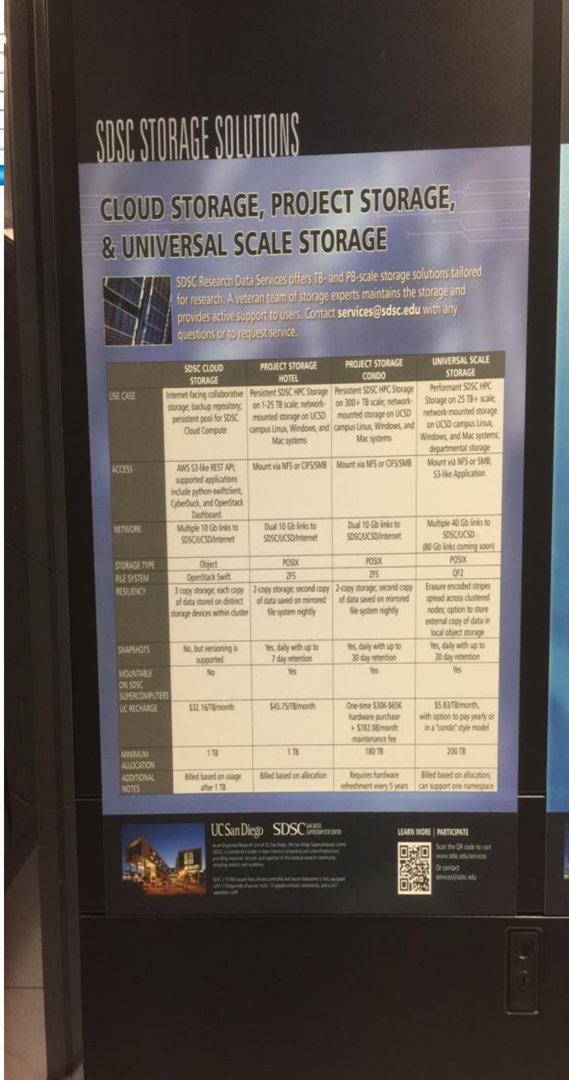
Issues & Recommendations:

- DTN user cases
- Prepare for 100G network data connectivity end to end tests
- DTN performance tuning over network

Dallas, TX | hpc inspires.

Connections


MREFC Projects
Commercial and Federal Clouds
HPC Centers XSEDE
OSN
Universities with 10-100G
University and Domain Repositories
Shared Backup for Disaster Recovery
Big Data Hubs



SDSC STORAGE SOLUTIONS

CLOUD STORAGE, PROJECT STORAGE, & UNIVERSAL SCALE STORAGE

SDSC Research Data Services offers TB- and PB-scale storage solutions tailored for research. A veteran team of storage experts maintains the storage and provides active support to users. Contact **services@sdsc.edu** with any questions or to request service.

	SDSC CLOUD STORAGE	PROJECT STORAGE HOTEL	PROJECT STORAGE CONDO	UNIVERSAL SCALE STORAGE
USE CASE	Internet-facing collaborative storage; backup repository; persistent pool for SDSC Cloud Compute	Persistent SDSC HPC Storage on 1-25 TB scale; network-mounted storage on UCSD campus Linux, Windows, and Mac systems	Persistent SDSC HPC Storage on 300+ TB scale; network-mounted storage on UCSD campus Linux, Windows, and Mac systems	Performant SDSC HPC Storage on 25 TB+ scale; network-mounted storage on UCSD campus Linux, Windows, and Mac systems; departmental storage
ACCESS	AWS S3-like REST API; supported applications include python-swiftclient, CyberDuck, and OpenStack Dashboard.	Mount via NFS or CIFS/SMB	Mount via NFS or CIFS/SMB	Mount via NFS or SMB; S3-like Application
NETWORK	Multiple 10 Gb links to SDSC/UCSD/Internet	Dual 10 Gb links to SDSC/UCSD/Internet	Dual 10 Gb links to SDSC/UCSD/Internet	Multiple 40 Gb links to SDSC/UCSD (80 Gb links coming soon)
STORAGE TYPE	Object	POSIX	POSIX	POSIX
FILE SYSTEM	OpenStack Swift	ZFS	ZFS	QF2
RESILIENCY	3 copy storage; each copy of data stored on distinct storage devices within cluster	2-copy storage; second copy of data saved on mirrored file system nightly	2-copy storage; second copy of data saved on mirrored file system nightly	Erasure encoded stripes spread across clustered nodes; option to store external copy of data in local object storage
SNAPSHOTS	No, but versioning is supported.	Yes, daily with up to 7 day retention	Yes, daily with up to 30 day retention	Yes, daily with up to 30 day retention
MOUNTABLE ON SDSC SUPERCOMPUTERS	No	Yes	Yes	Yes
UC RECHARGE	$32.16/TB/month	$45.75/TB/month	One-time $30K-$65K hardware purchase + $782.08/month maintenance fee	$5.83/TB/month, with option to pay yearly or in a "condo" style model
MINIMUM ALLOCATION	1 TB	1 TB	180 TB	200 TB
ADDITIONAL NOTES	Billed based on usage after 1 TB	Billed based on allocation	Requires hardware refreshment every 5 years	Billed based on allocation; can support one namespace

UC San Diego SDSC SAN DIEGO SUPERCOMPUTER CENTER

As an Organized Research Unit of UC San Diego, the San Diego Supercomputer Center (SDSC) is considered a leader in data-intensive computing and cyberinfrastructure, providing resources, services, and expertise to the national research community, including industry and academia.

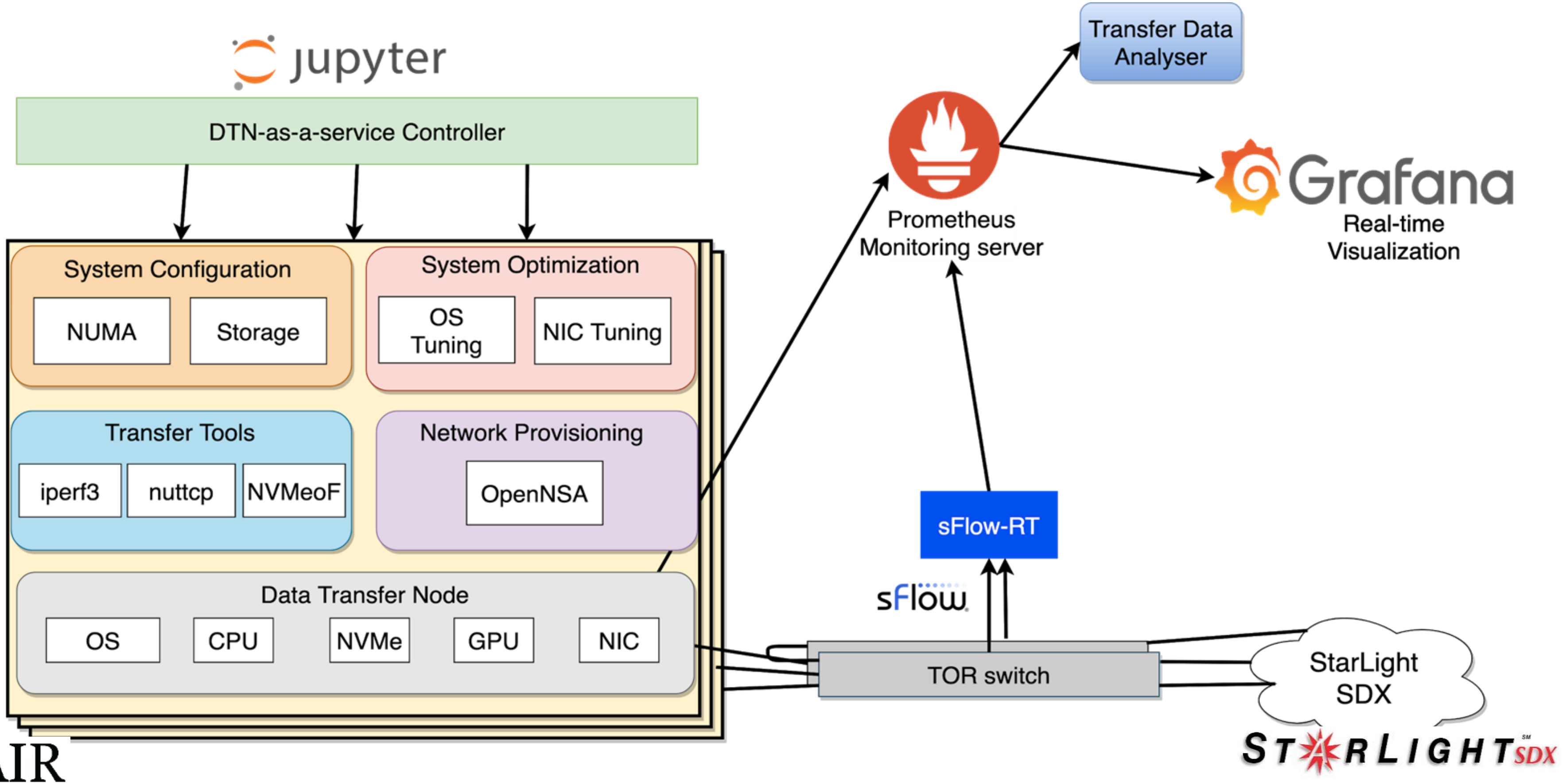
SDSC's 19,000 square foot climate-controlled and secure datacenter is fully equipped with 13 Megawatts of power, multi- 10 gigabit network connectivity, and a 24/7 operations staff.

LEARN MORE | PARTICIPATE

Scan the QR code to visit www.sdsc.edu/services

Or contact services@sdsc.edu

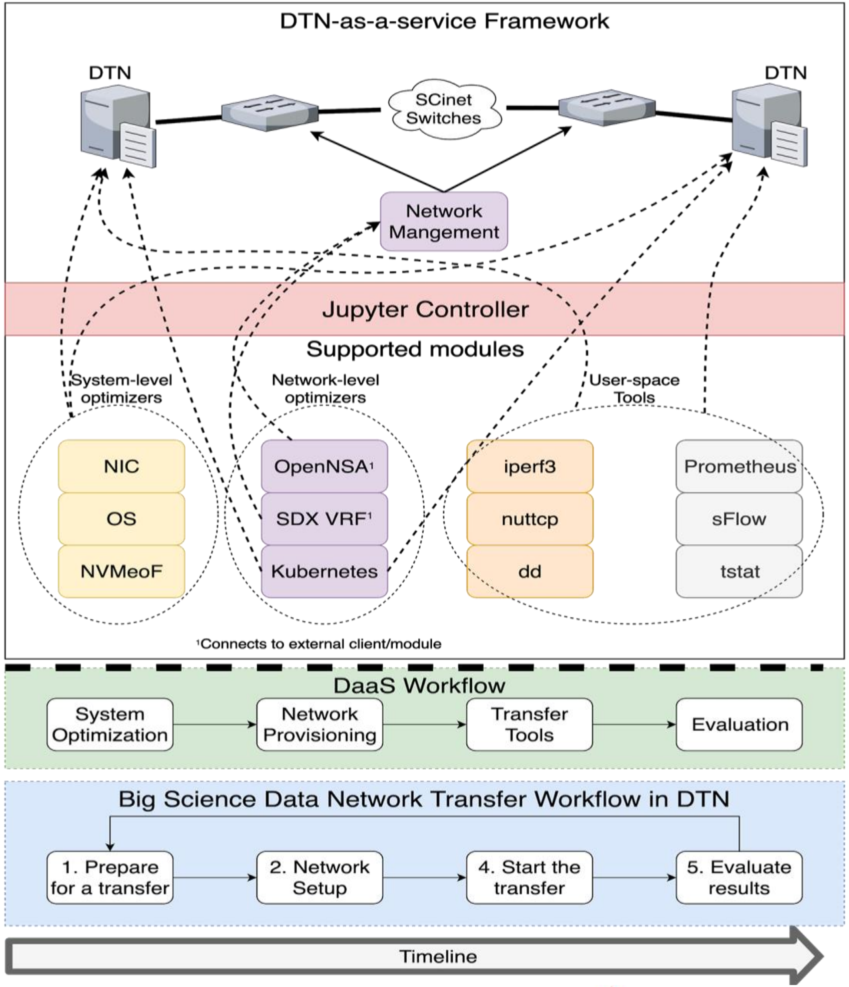
Current Starlight DTN-as-a-Service Software Stack Architecture



iCAIR

Mapping DTNaaS to Big Science data transfer workflow

- DTNaaS workflow maps Big Science data transfer workflow with DTN
- Each module corresponds to procedures for data transfer
- Jupyter Controller implements the workflow integration
- Transfer monitoring and evaluation provides analysis for the workflow



Managing Resources

Each module manages the following resources of DTN

(Host machine-specific in **bold, require sudo**)

System Configuration module

- CPU type and NUMA node information
- Available service ports

System Optimization module

- **TCP/IP stack parameters**
- **NIC parameters**
- **Linux traffic control parameters**
- **PCIe connection parameters**
- **CPU type-specific parameters**

Monitoring resources

Transfer tools module

- Available transfer protocols: NUTTCP, NVMeoF

DTN Monitoring (node_exporter)

- Physical hardware : CPU, SATA, NVMe, Memory
- Network : infiniband, netdev, ARP, IPVS, sockstat
- Disks : filesystem, diskstats, ZFS, XFS
- OS : vmstat, stat, hwmon

Network Monitoring (sflow)

- Port counters : Errors, Collisions, Discards, octets, packets, utilization, broadcast, speed
- Protocol specific counters : ARP, DHCP, DNS, ICMP, IP, LLDP, NTP, TCP, UDP, VLAN

Starlight DTNaaS Software Stack

Optimize the transfer performance based on the system configuration

Provide functions to automate data transfer

Set up and tear down transfer-tool environment on the DTNs

Modular component to support additional data transfer tools

- Provided system configuration and optimization module
- iperf3, nuttcp, and NVMeoF for transferring data in high-speed
- Workflow controller implemented in Jupyter to enable easy research & collaboration

Tuning on Jupyter

Tuning Units

-irqbalance off

-Increase TCP buffer to 2GB

-Fair Queuing : Pacing inter-packet gap

-MTU: Jumbo frames

-CPU_gorvernor: Performance mode

-Ring_buffer : NIC ring buffer to 8k

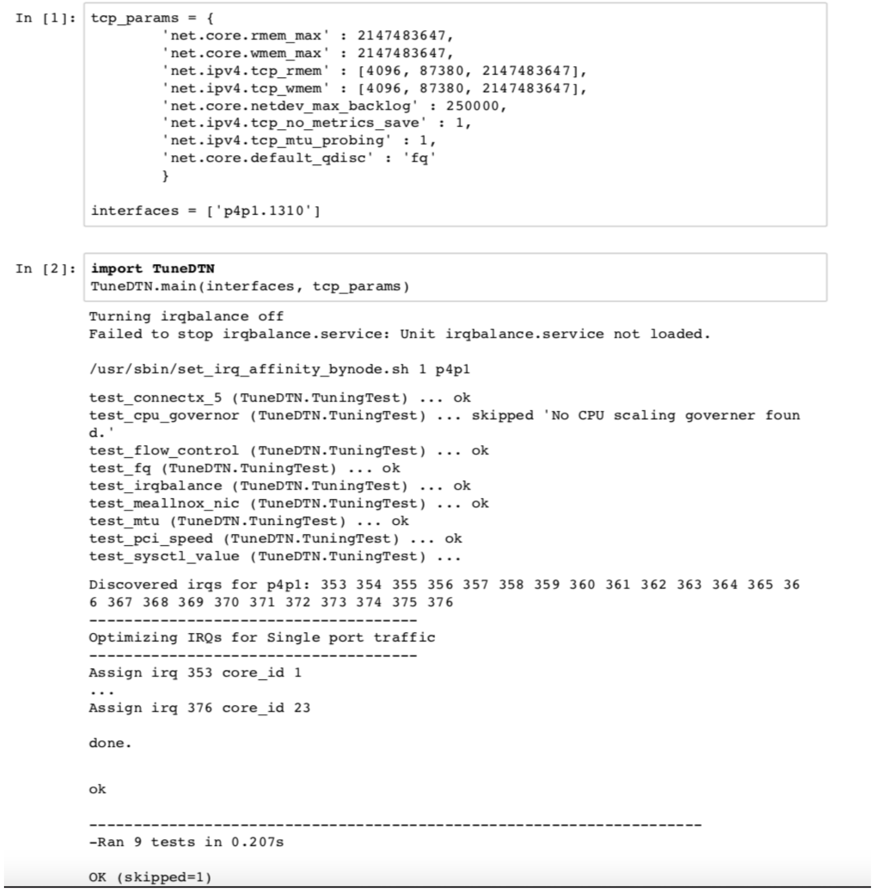
-Ethernet Flow Control: On

-Bind NIC irq to the local NUMA node

*Mellanox 100G NIC specific tuning

- Set PCIe Maxreadreq to 4096

*AMD specific tuning



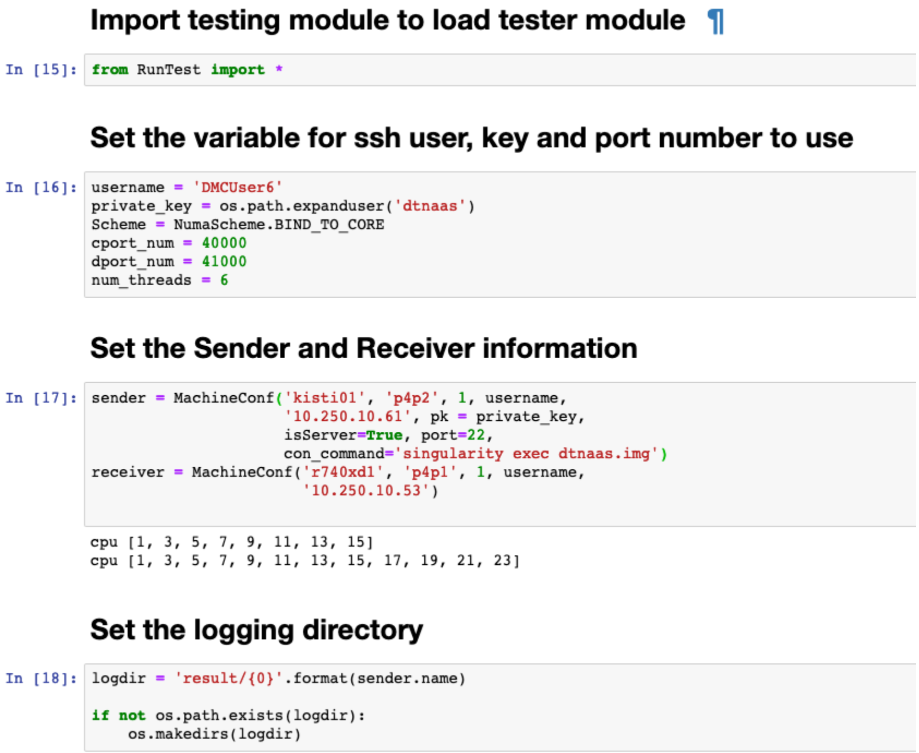
```
In [1]: tcp_params = {
            'net.core.rmem_max' : 2147483647,
            'net.core.wmem_max' : 2147483647,
            'net.ipv4.tcp_rmem' : [4096, 87380, 2147483647],
            'net.ipv4.tcp_wmem' : [4096, 87380, 2147483647],
            'net.core.netdev_max_backlog' : 250000,
            'net.ipv4.tcp_no_metrics_save' : 1,
            'net.ipv4.tcp_mtu_probing' : 1,
            'net.core.default_qdisc' : 'fq'
            }

        interfaces = ['p4p1.1310']
```

```
In [2]: import TuneDTN
        TuneDTN.main(interfaces, tcp_params)
```

```
Turning irqbalance off
Failed to stop irqbalance.service: Unit irqbalance.service not loaded.

/usr/sbin/set_irq_affinity_bynode.sh 1 p4p1

test_connectx_5 (TuneDTN.TuningTest) ... ok
test_cpu_governor (TuneDTN.TuningTest) ... skipped 'No CPU scaling governer foun
d.'
test_flow_control (TuneDTN.TuningTest) ... ok
test_fq (TuneDTN.TuningTest) ... ok
test_irqbalance (TuneDTN.TuningTest) ... ok
test_meallnox_nic (TuneDTN.TuningTest) ... ok
test_mtu (TuneDTN.TuningTest) ... ok
test_pci_speed (TuneDTN.TuningTest) ... ok
test_sysctl_value (TuneDTN.TuningTest) ...

Discovered irqs for p4p1: 353 354 355 356 357 358 359 360 361 362 363 364 365 36
6 367 368 369 370 371 372 373 374 375 376
-------------------------------------
Optimizing IRQs for Single port traffic
-------------------------------------
Assign irq 353 core_id 1
...
Assign irq 376 core_id 23

done.


ok

----------------------------------------------------------------------
-Ran 9 tests in 0.207s

OK (skipped=1)
```

Run transfers on Jupyter

Step 1: Follow Jupyter notebook to set-up the DTNs



Import testing module to load tester module ¶

```
In [15]: from RunTest import *
```

Set the variable for ssh user, key and port number to use

```
In [16]: username = 'DMCUser6'
         private_key = os.path.expanduser('dtnaas')
         Scheme = NumaScheme.BIND_TO_CORE
         cport_num = 40000
         dport_num = 41000
         num_threads = 6
```

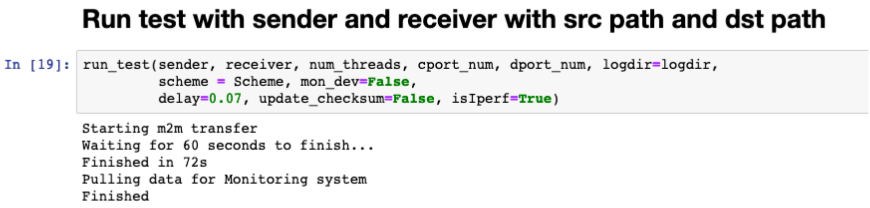
Set the Sender and Receiver information

```
In [17]: sender = MachineConf('kisti01', 'p4p2', 1, username,
                              '10.250.10.61', pk = private_key,
                              isServer=True, port=22,
                              con_command='singularity exec dtnaas.img')
         receiver = MachineConf('r740xd1', 'p4p1', 1, username,
                                '10.250.10.53')

         cpu [1, 3, 5, 7, 9, 11, 13, 15]
         cpu [1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23]
```

Set the logging directory

```
In [18]: logdir = 'result/{0}'.format(sender.name)

         if not os.path.exists(logdir):
             os.makedirs(logdir)
```

Step 2: Specify the type of action and run

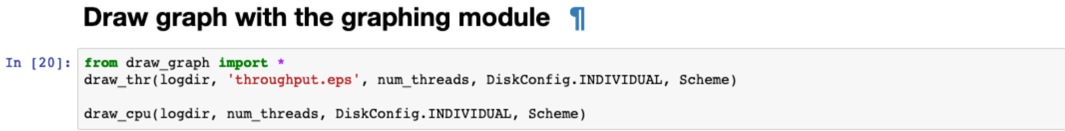


Run test with sender and receiver with src path and dst path

```
In [19]: run_test(sender, receiver, num_threads, cport_num, dport_num, logdir=logdir,
                  scheme = Scheme, mon_dev=False,
                  delay=0.07, update_checksum=False, isIperf=True)

         Starting m2m transfer
         Waiting for 60 seconds to finish...
         Finished in 72s
         Pulling data for Monitoring system
         Finished
```

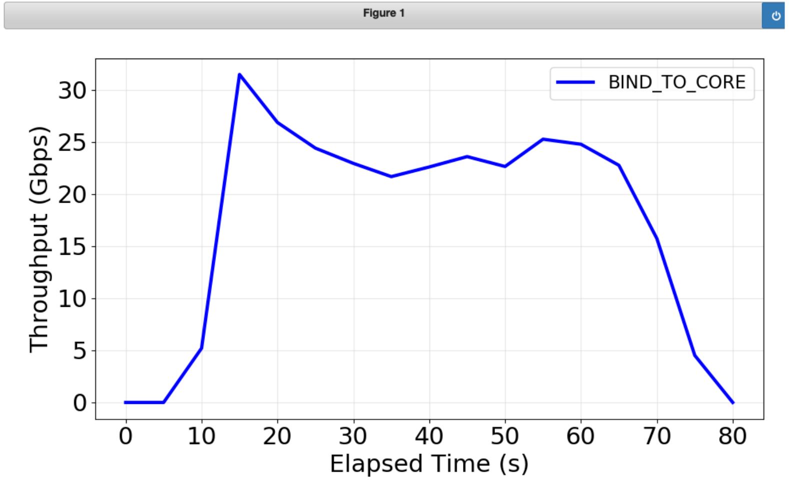
Run file transfer on a Jupyter notebook

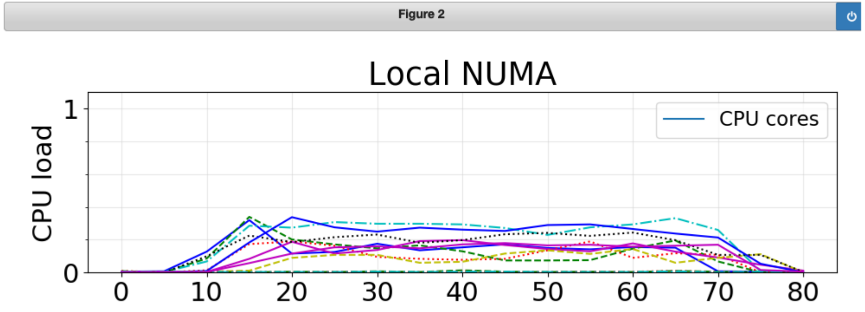
Step 3: Draw the Graph!



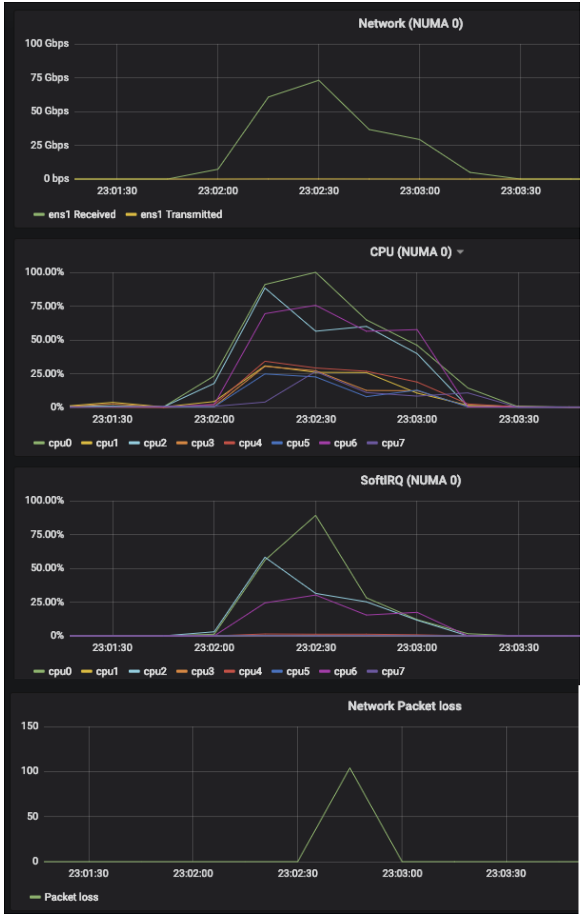
Draw graph with the graphing module ¶

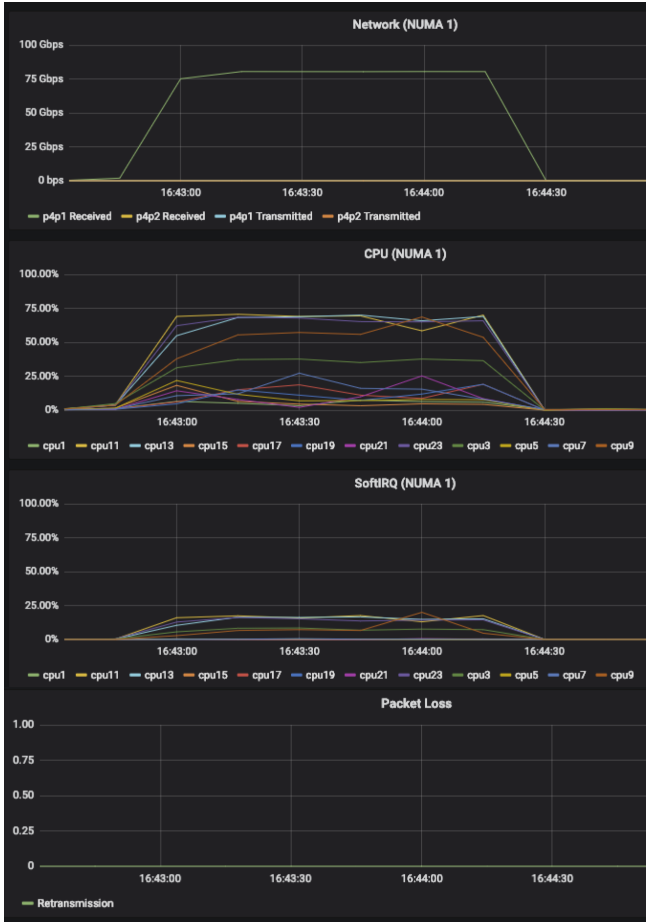
```
In [20]: from draw_graph import *
         draw_thr(logdir, 'throughput.eps', num_threads, DiskConfig.INDIVIDUAL, Scheme)

         draw_cpu(logdir, num_threads, DiskConfig.INDIVIDUAL, Scheme)
```

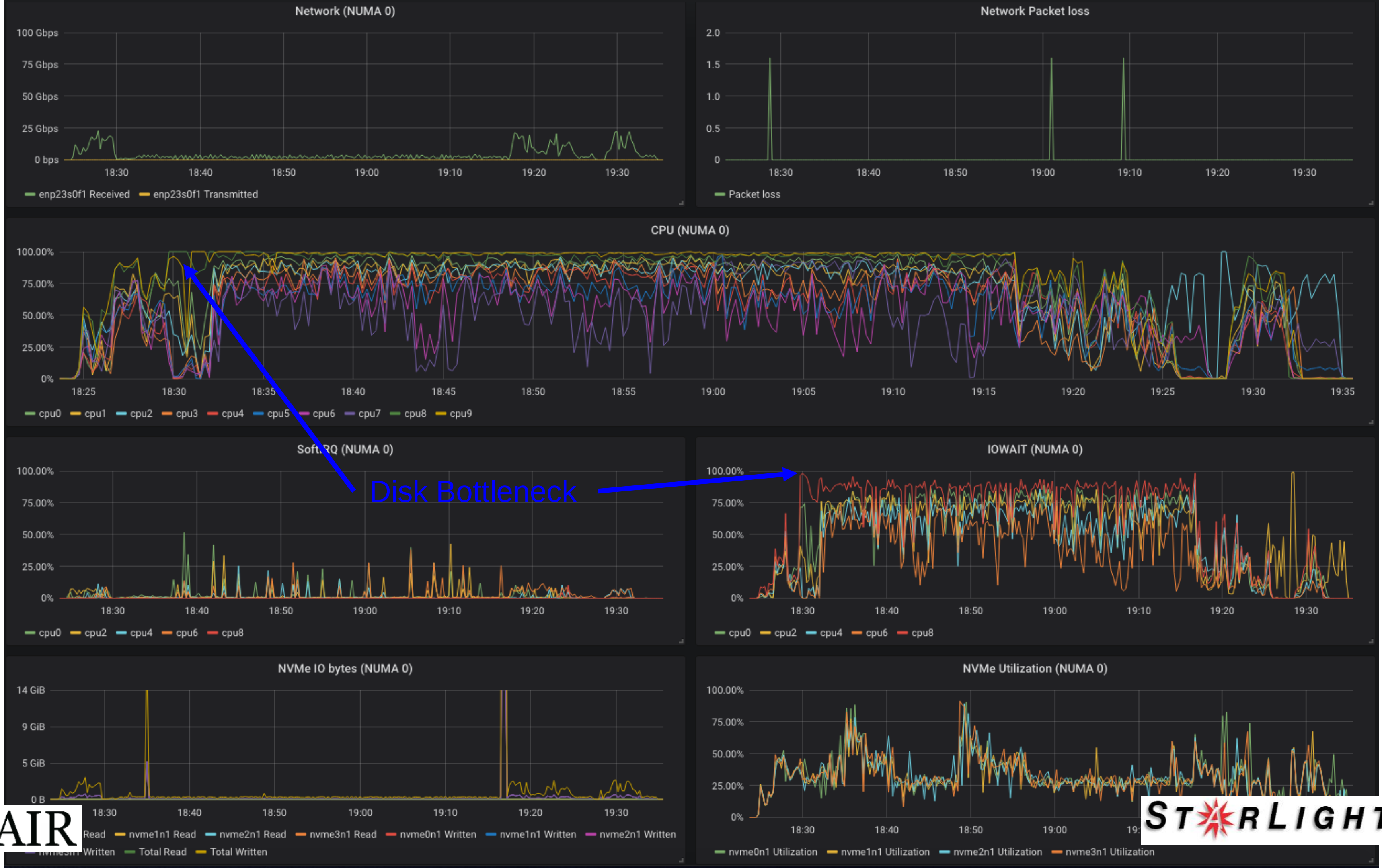




Transmission with/without packet loss







Disk Bottleneck

iCAIR

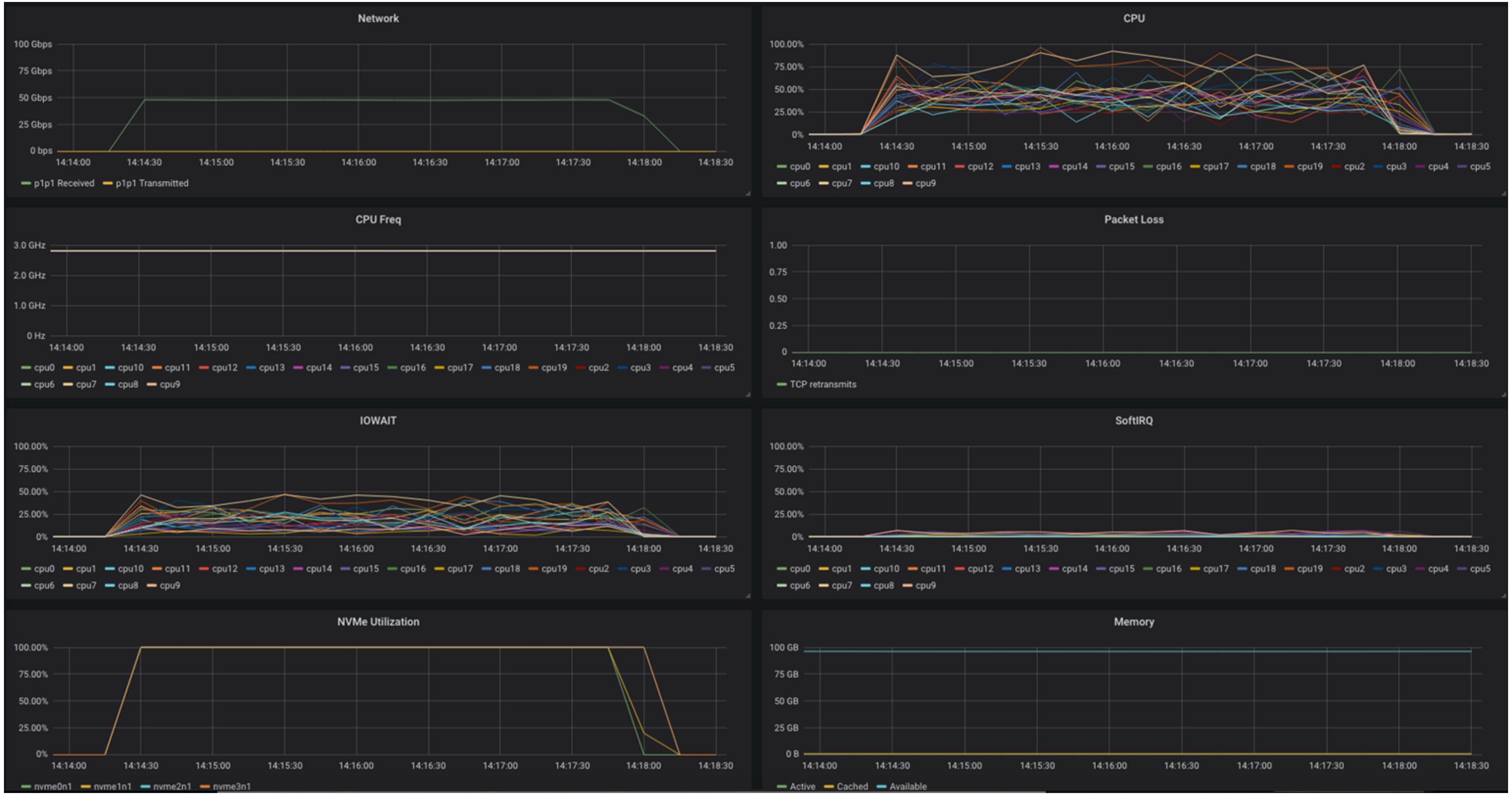
STARLIGHT SDX

NVMe over Fabrics

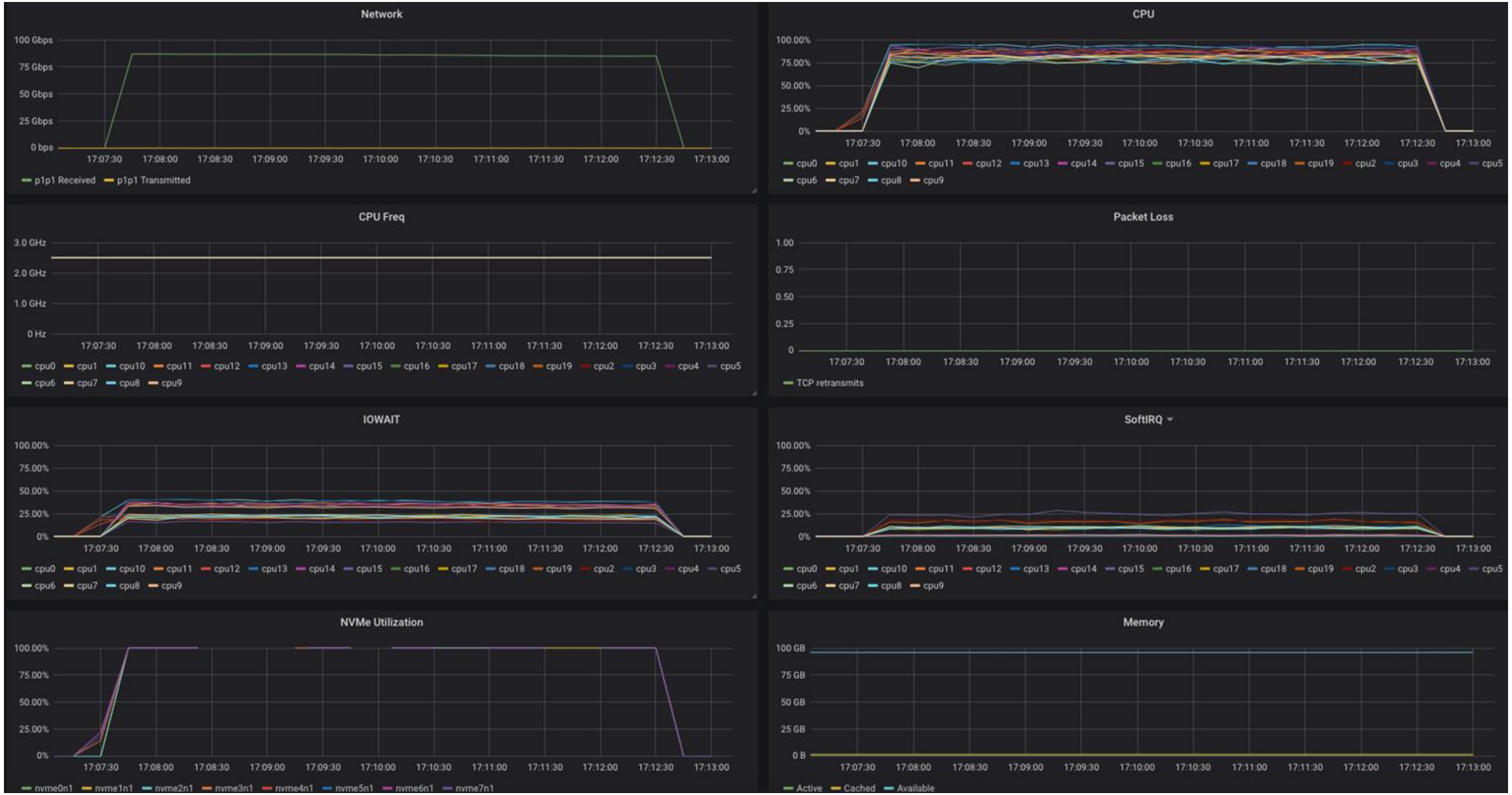
NVMe over Fabrics features:

- Accessing remote NVMe device over LAN or WAN
- RDMA and TCP fabrics support
- Allow for instance data access
- Suitable for streaming data or remote data access
- Low overhead
- Efficient

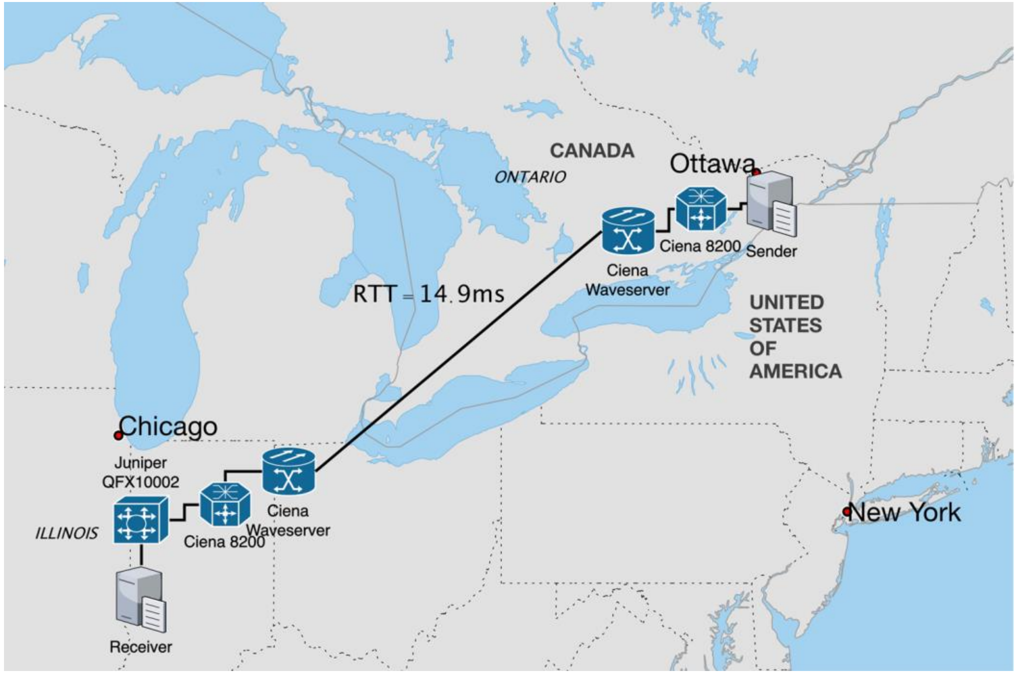
NVMe transfer with one NVMe x8 card in LAN



NVMe transfer with two NVMe x8 card in LAN

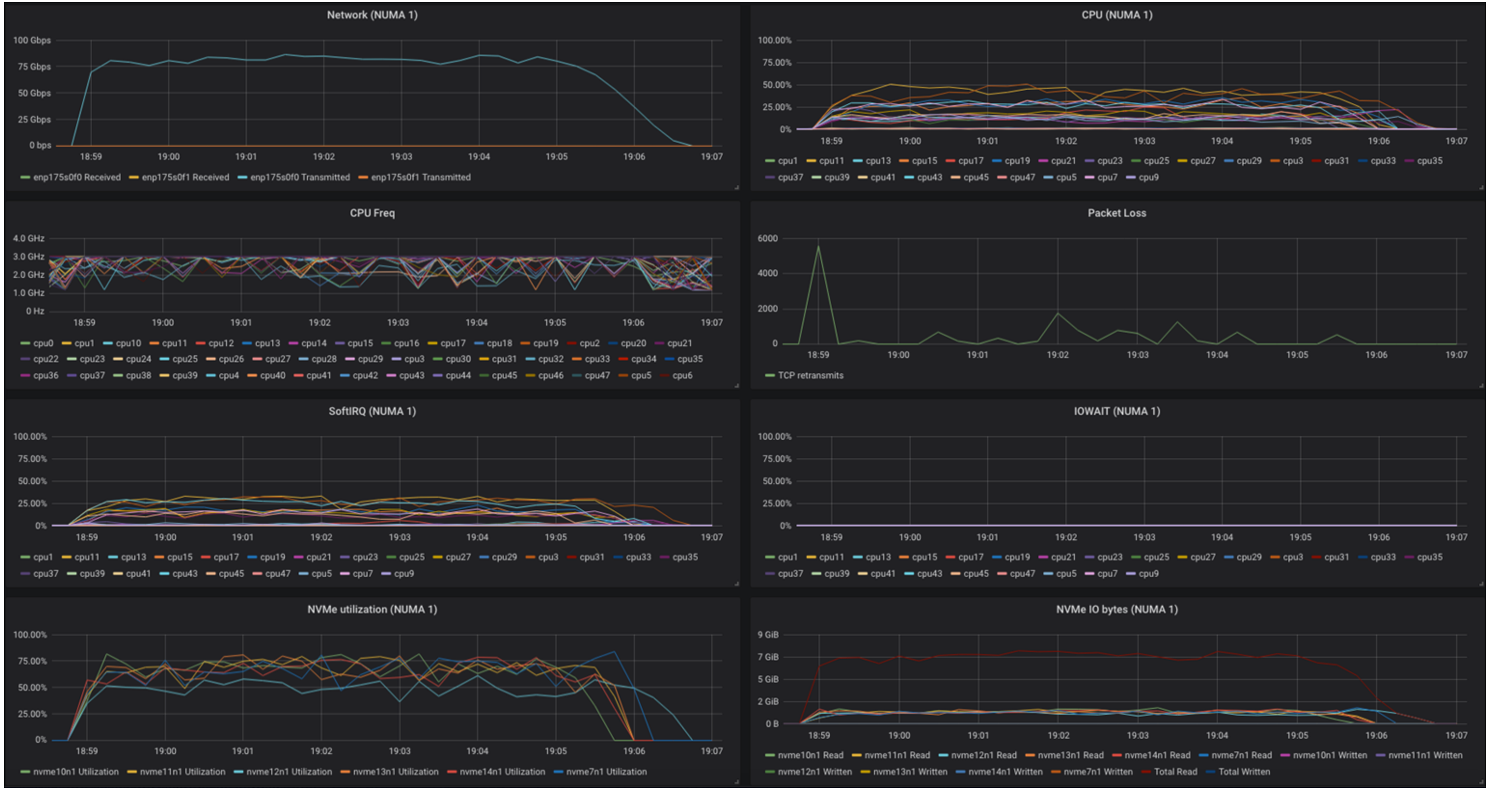


NVMe over Fabrics with TCP over a long distance

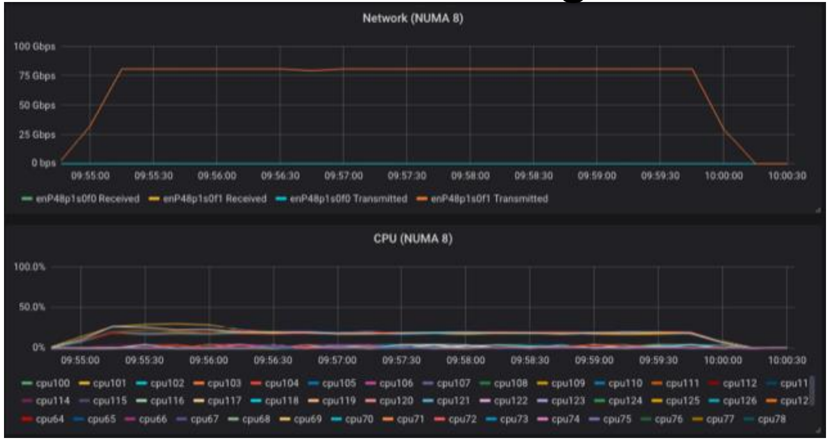


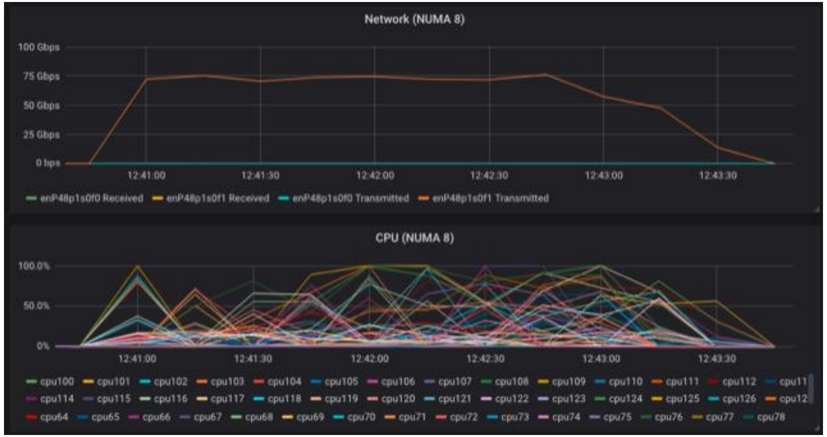
	Sender	Receiver
CPU	2 * Intel(R) Xeon(R) Gold 6136 CPU @ 3.00GHz	
Memory	DDR4-2666 192 GB	
NIC	Mellanox Technologies MT27800 Family [ConnectX-5]	
NVME	2 * Kingston DCP1000 (4 * 800 GB each)	8 * Samsung SSD 960 PRO 2TB
OS	GNU/Linux 5.1.0.rc4	
File System	XFS	XFS

NVMe over Fabrics with TCP over a long distance

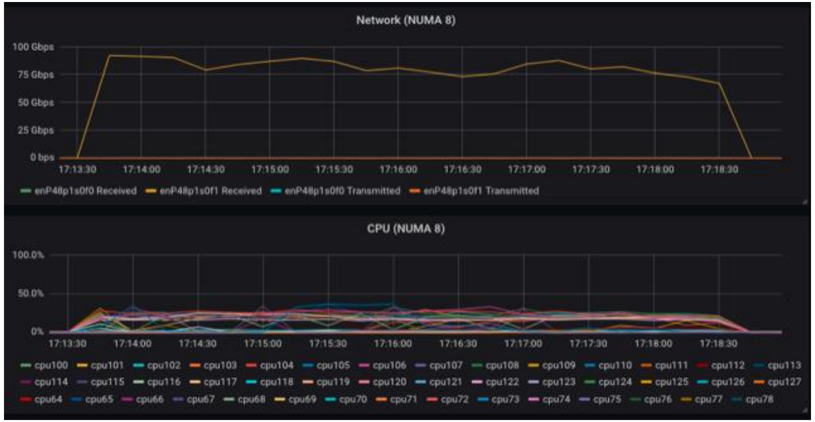


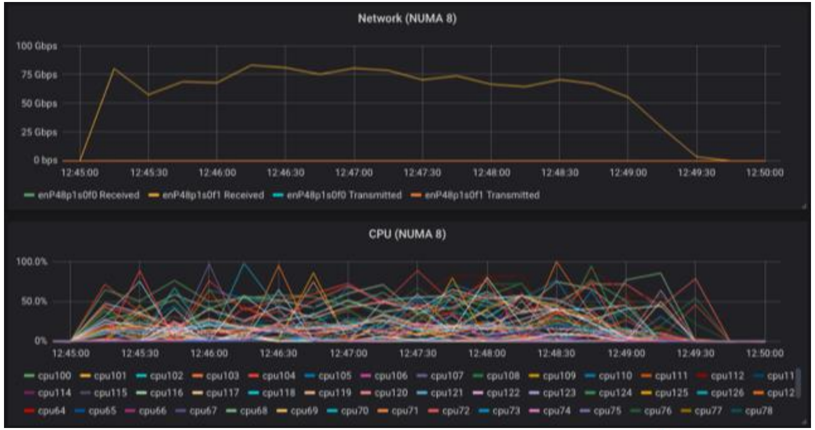
CERN-Starlight





Starlight-CERN





iCAIR

STARLIGHT SDX

SDX DTNaaS Future Work

- OSG DTNaaS prototype and national and international OSG Cache DTNaaS trial(Summary from 2nd OSG-IRNC workshop, Sep 16 2019)
- Partner with big data science community and regional/national/international SDXs to establish LAN/WAN packet loss trouble shooting reference workflow and procedure
- XrootD and other protocol integration prototype
- SDX NVMeoF Service Prototype
- DTNaaS clustering and federation prototype