Pacific Research Platform: Building an Accelerator for Discovery

John Hess
CENIC - Pacific Wave

Pacific Research Platform: Background

Created in 2015, the Pacific Research Platform (PRP) is an NSF-funded, regional project to meet the needs of researchers in California, the Western U.S., and beyond. CENIC and Pacific Wave provide the PRP with a high-speed backplane and connectivity within the region and to other R&E networks.

The PRP is a project to accelerate discovery:

- by improving end-to-end, high-speed data transfer, data placement and storage, and distributed computing capabilities in collaborative, big-data science; and,

- by orchestrating direct engagements between researchers and cyber-infrastructure engineers to assess and optimize scientific workflows, to inform technical requirements, and drive project priorities.

The PRP's scope covers a broad range of data-intensive research including projects from the fields of particle physics, astronomy, biomedical sciences, earth sciences, and computer science and engineering.

The PRP is a partnership of more than 20 institutions, including four national supercomputer centers.

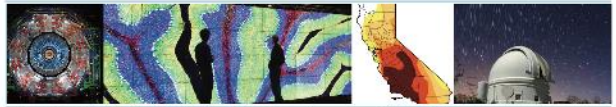
Pacific Research Platform: Overview

PRP PACIFIC RESEARCH PLATFORM

Pacific Research Platform: The Future of Big Data Collaboration

From biomedical sciences to particle physics, today nearly all research and data analysis involves remote collaboration. In order to work effectively and efficiently on multi-institutional projects, researchers depend heavily on high-speed access to large datasets and computing resources.

To meet the needs of researchers in California and beyond, the National Science Foundation (NSF) has awarded a five-year, $5 million grant to fund the Pacific Research Platform (PRP). The PRP integrates Science DMZs, an architecture developed by the U.S. Department of Energy's Energy Sciences Network (ESnet), into a high-capacity regional "freeway system." This system makes it possible for large amounts of scientific data to be moved between scientists' labs and their collaborators' sites, supercomputer centers or data repositories, without performance degradation.



A Regional Model for Multi-Discipline Data-Intensive Networking

The PRP, led by researchers at UC San Diego and UC Berkeley, will enable fast and secure data transfers between researchers in over 20 universities. The PRP builds on the optical backbone of Pacific Wave, a joint project of CENIC and the Pacific Northwest GigaPOP (PNWGP) to create one large, seamless research platform that will encourage statewide, regional—even worldwide—collaboration.

The PRP will support a broad range of data-intensive research projects that will have wide-reaching impacts on science and technology worldwide. Cancer genomics, galaxy evolution research, climate modeling, and the creation of virtual reality gaming systems are just a few of the projects that will benefit from the PRP.

Principal Investigator

Larry Smarr
UC San Diego (UCSD), California Institute for Telecommunications and Information Technology (Calit2)

Co-Principal Investigators

Camille Crittenden
UC Berkeley, Center for Information Technology Research in the Interest of Society (CITRIS) and the Banatao Institute

Tom DeFanti
UCSD, Calit2/Qualcomm Institute

Philip Papadopoulos
UCSD, San Diego Supercomputer Center

Frank Wuerthwein
Physics Dept. and UCSD, San Diego Supercomputer Center

During a demonstration of its capabilities at the 2015 CENIC Conference, researchers showed that **the PRP moved data up to 500 times faster** *than speeds currently available.*

Pacific Research Platform: Cyberinfrastructure for Big Data

The PRP's data-sharing architecture, with disk-to-disk 10-100Gbps connections, enables region-wide virtual co-location of data with computing resources. Today, dozens of top universities and research centers are doing work across ten major application areas, positioning the PRP to be a regional-scale model for a future national-scale Big Data cyberinfrastructure.

https://cenic.org/files/publications/PRP_Update_2016_F.pdf

Agenda

This session will discuss the PRP architecture and operations over CENIC and Pacific Wave, the technical challenges faced and successes realized by the PRP during its first three years of activities, explore the emerging near-term and longer-term PRP capabilities, plans for expansion, as well as an invitation to collaborate as we explore scaling from a regional network to national and international models:

PRP: Extending the ESnet Science DMZ model to a regional scale

- o PRP: CENIC + Pacific Wave, and partner R&E networks
- o DTN architecture
- o Network and disk-to-disk performance measurement

Science Engagement & Case Studies -- Socio-technical engineering among science, HPC, network, and IT, with a selection of PRP case-studies driving the direction of the PRP

PRP dev: A testbed incorporating security, IPv6, SDN/SDX, cooperating research groups

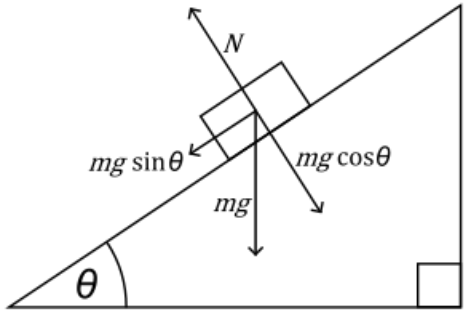
- o Nautilus: PRP implementation of Kubernetes -- virtualized, federated access to distributed cluster resources including Rook-orchestrated Ceph for persistent storage and GPU-computing.
- o Super-Channels / Super-Facilities
- o Interdomain collaborations

Considerations -- scaling to national and international models

The Science DMZ* in 1 Slide

Consists of **three key components**, all required:

- "Friction free" network path
  - Highly capable network devices (wire-speed, deep queues)
  - Virtual circuit connectivity option
  - Security policy and enforcement specific to science workflows
  - Located at or near site perimeter if possible
- Dedicated, high-performance Data Transfer Nodes (DTNs)
  - Hardware, operating system, libraries all optimized for transfer
  - Includes optimized data transfer tools such as Globus Online and GridFTP
- Performance measurement/test node
  - perfSONAR
- Engagement with end users



globus

perfSONAR

Details at

http://fasterdata.es.net/science-dmz/

* ***Science DMZ*** is a trademark of The Energy Sciences Network (ESnet)

ESnet

PRP's First 2 Years: Connecting Multi-Campus Application Teams and Devices

Particle Physics

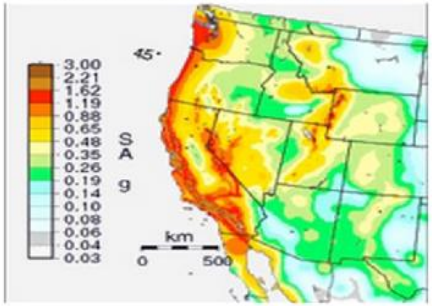
Biomedical 'omics

CENIC

ESnet

Telescope Surveys

Earthquake Engineering



NSF

Visualization, Virtual Reality, Collaboration

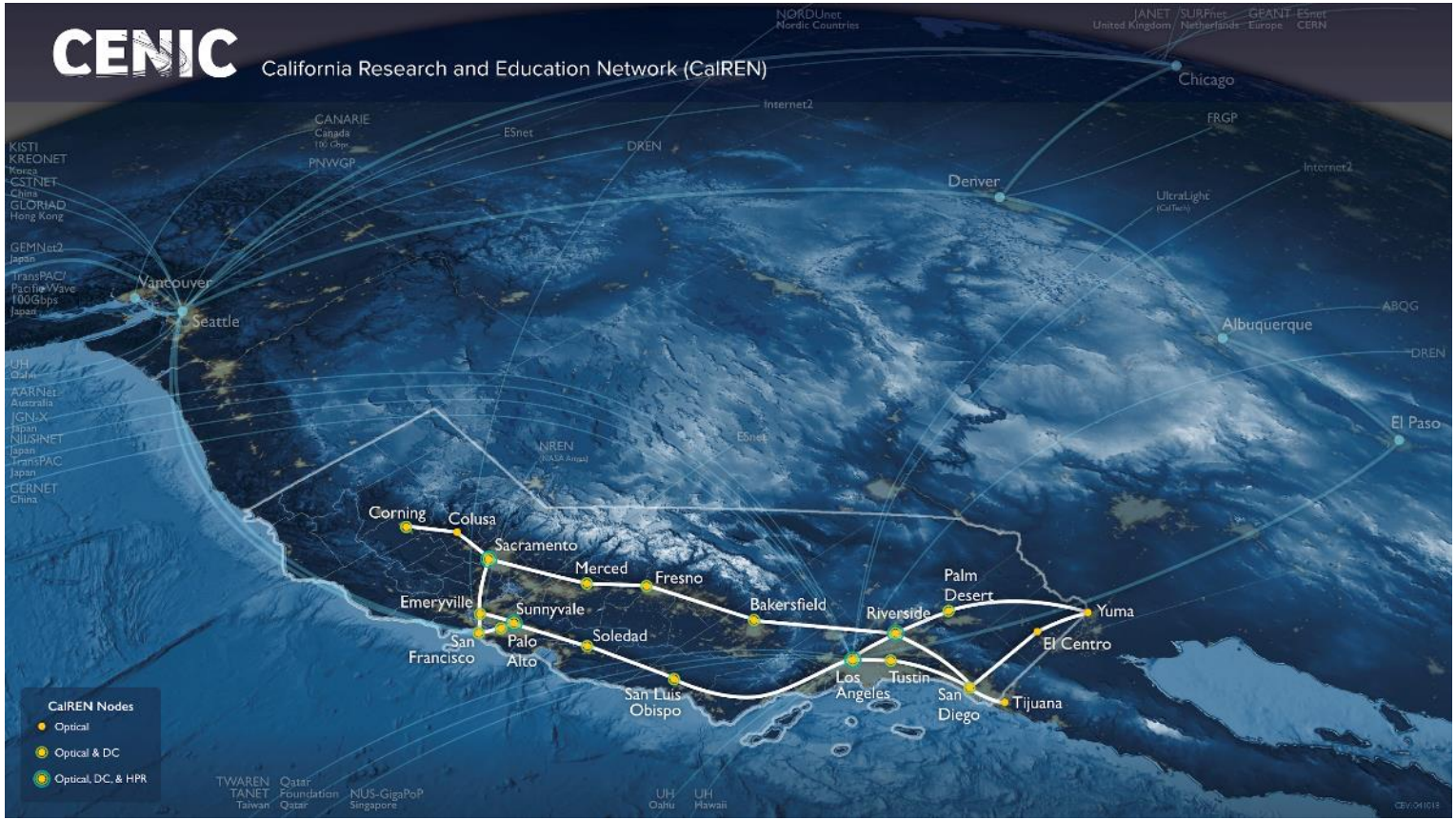
CITRIS AND THE BANATAO INSTITUTE

it²

PRP

CENIC

Pacific Research Platform: CENIC CalREN



https://cenic.org/files/publications/Pacific_Wave_Map_2017.jpg

Pacific Research Platform: Pacific Wave distributed exchange

https://cenic.org/files/publications/CENIC_CalREN_4-10-18.001.pdf

FIONAs and FIONettes

- FIONAs are USD $8,000 PCs [a.k.a ESnet Data Transfer Nodes (DTNs)] with:
  - 10/40Gbps or 100Gbps network interface cards
  - 3TB Solid state disks (SSDs) to achieve nearly 40Gbps disk-to-disk
  - +NVMe SSDs to achieve nearly 100Gbps disk-to-disk
  - +Up to 8 Nvidia or AMD GPUs for Machine Learning
  - +Up to 16 10TB internal hard disk drives for data posting
  - +Up to 38 Intel or AMD CPU cores; more RAM for computation
- USD $750 10Gbps FIONAs are being tested; add RAM, SSDs, 10TB drives to suit.
- FIONettes are USD $250 1Gbps Gigabyte EL-20 4-core Pentium IoT gateways
  - Used in hands-on workshops at CENIC2018 (3-4 March: Monterey), NORDUnet - SURFnet (16-17 April: Copenhagen), GPN (30 May - 1 June: Kansas City), and Montana State University (2-5 August: Bozeman)
  - Workshop participants will build them, configure them to register iperf3 (memory-to-memory) and GridFTP (disk-to-disk) throughput results to a central Esmond Measurement Archive, setup grids with MaDDash to visualize results, learn to identify and correct network pathologies routes, maintain them, and then take them home.

Source: Larry Smarr, Calit2

PRP Science DMZ Data Transfer Nodes (DTNs) - Flash I/O Network Appliances (FIONAs)

UCSD Designed FIONAs To Solve the Disk-to-Disk Data Transfer Problem *at Full Speed* on 10G, 40G and 100G Networks

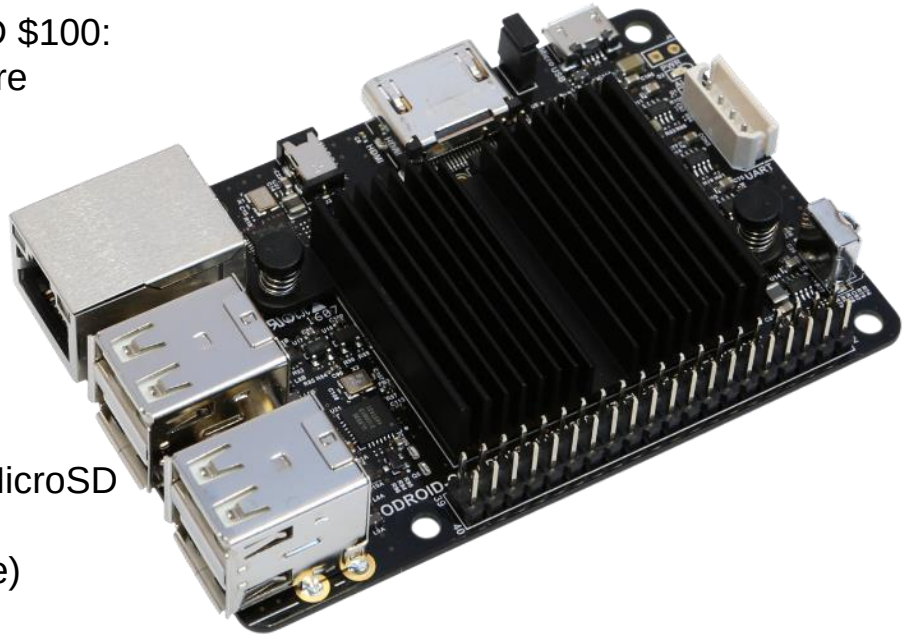
FIONette v2 -- USD $250
Gigabyte EL-20-3700-32GB
Intel® Pentium® processor N3710
2 x SO-DIMM DDR3L slots
2 x GbE LAN ports
32GB onboard eMMC memory
1 x Full-size Mini-PCIe slot

Image source: Gigabyte

Sources: Phil Papadopoulos, SDSC &
Tom DeFanti, Joe Keefe & John Graham, Calit2

Low-cost perfSONAR nodes

ODROID-C2 - 1Gbps, 2GB RAM, 32GB eMMC5.0: USD $100:
* Amlogic ARM® Cortex®-A53(ARMv8) 1.5Ghz quad core CPUs
* Mali™-450 GPU (3 Pixel-processors + 2 Vertex shader processors)
* 2Gbyte DDR3 SDRAM
* Gigabit Ethernet
* HDMI 2.0 4K/60Hz display
* H.265 4K/60FPS and H.264 4K/30FPS capable VPU
* 40pin GPIOs + 7pin I2S
* eMMC5.0 HS400 Flash Storage slot / UHS-1 SDR50 MicroSD Card slot
* USB 2.0 Host x 4, USB OTG x 1 (power + data capable)



Project tracking candidate, low-cost hardware:
https://github.com/perfsonar/project/wiki/Low-Cost-perfSONAR-Nodes

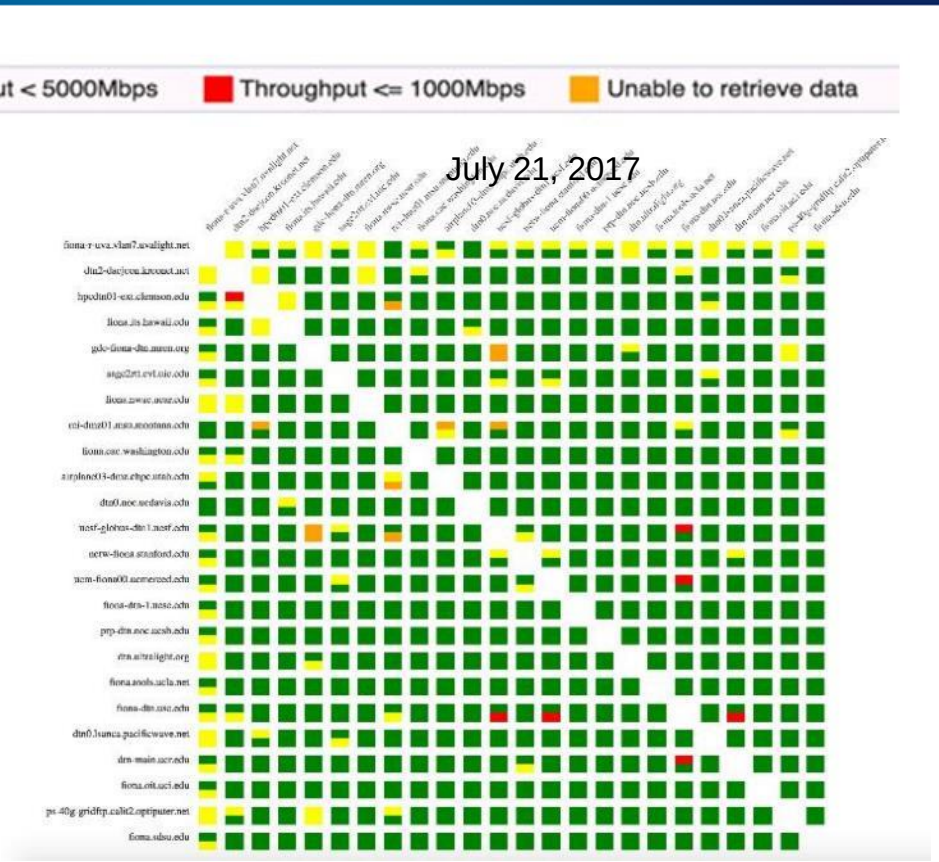
FIONAs on the PRP and Partners

- ~24 FIONAs are on the PRP as GridFTP (MaDDash) + perfSONAR nodes
  - PRP partners: 9 UCs, Caltech, USC, Stanford, SDSC, UW, UIC/EVL
  - Plus U Utah, Montana State, U Chicago, Clemson U, NCAR, U Hawaii, U Guam
  - Plus internationals: Uv Amsterdam, KISTI (Korea)
- Many states and regionals building FIONAs and creating MaDDashes
  - FIONA build specs on PRP website—several sources, new builds coming
  - FIONette workshop agenda and support materials will be posted
  - Weekly engineering calls with notes going to 140+ technical

More requests for FIONA workshops than we can handle: Indiana U/APAN, GPN, LEARN (TX)

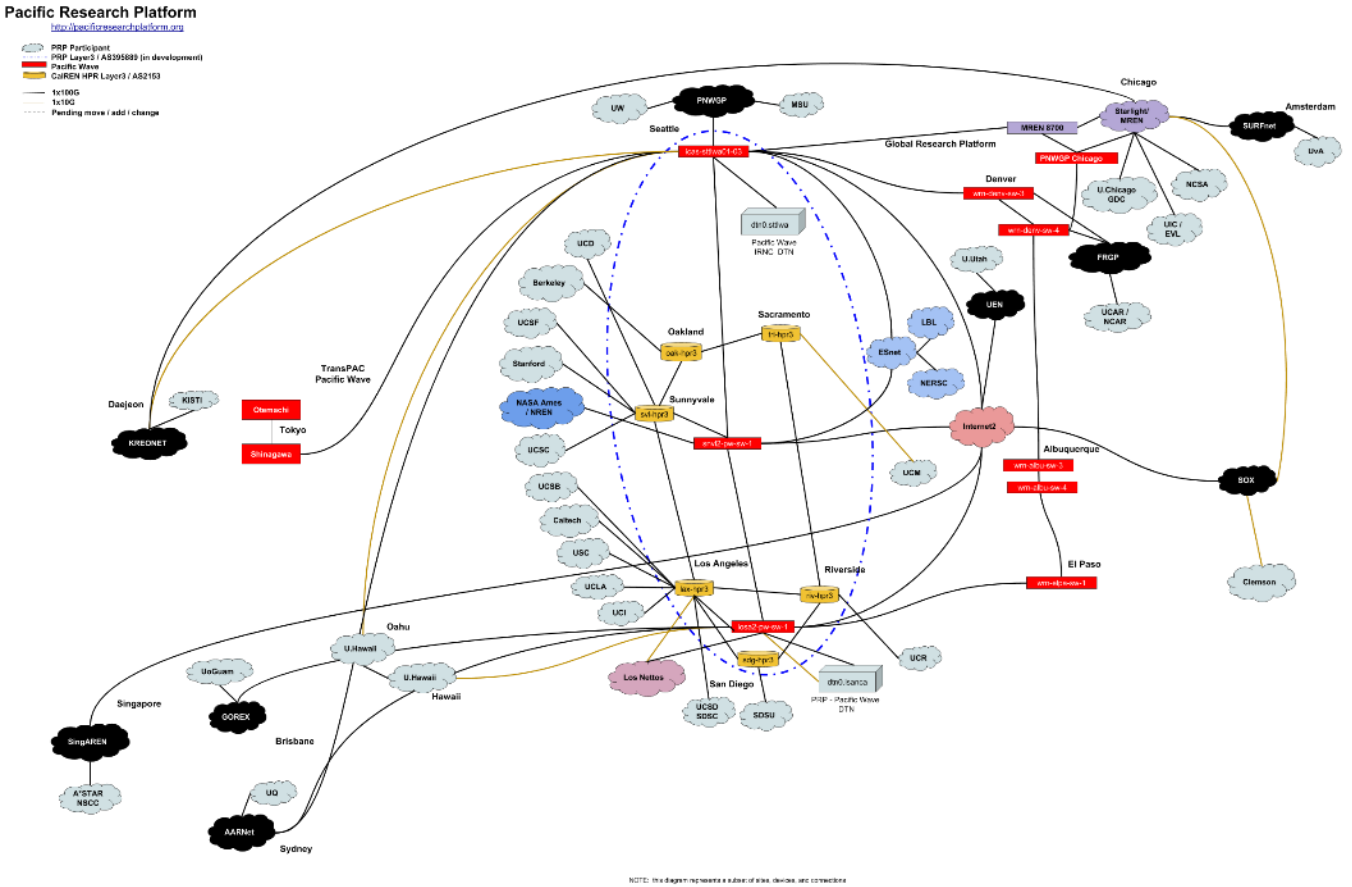
Source: Larry Smarr, Calit2

We Measured Disk-to-Disk Throughput with 10GB File Transfer 4 Times Per Day in Both Directions for All PRP Sites



Source: John Graham, Calit2

Pacific Research Platform: high-level topology diagram



https://docs.google.com/drawings/d/158wd7pAWg5mUnyppJJHxGBOxP-H0UJmvTRQ70V2sGJs/edit?usp=sharing

New NSF CHASE-CI Grant Creates a Community Cyberinfrastructure: Adding a Machine Learning Layer on top of the PRP

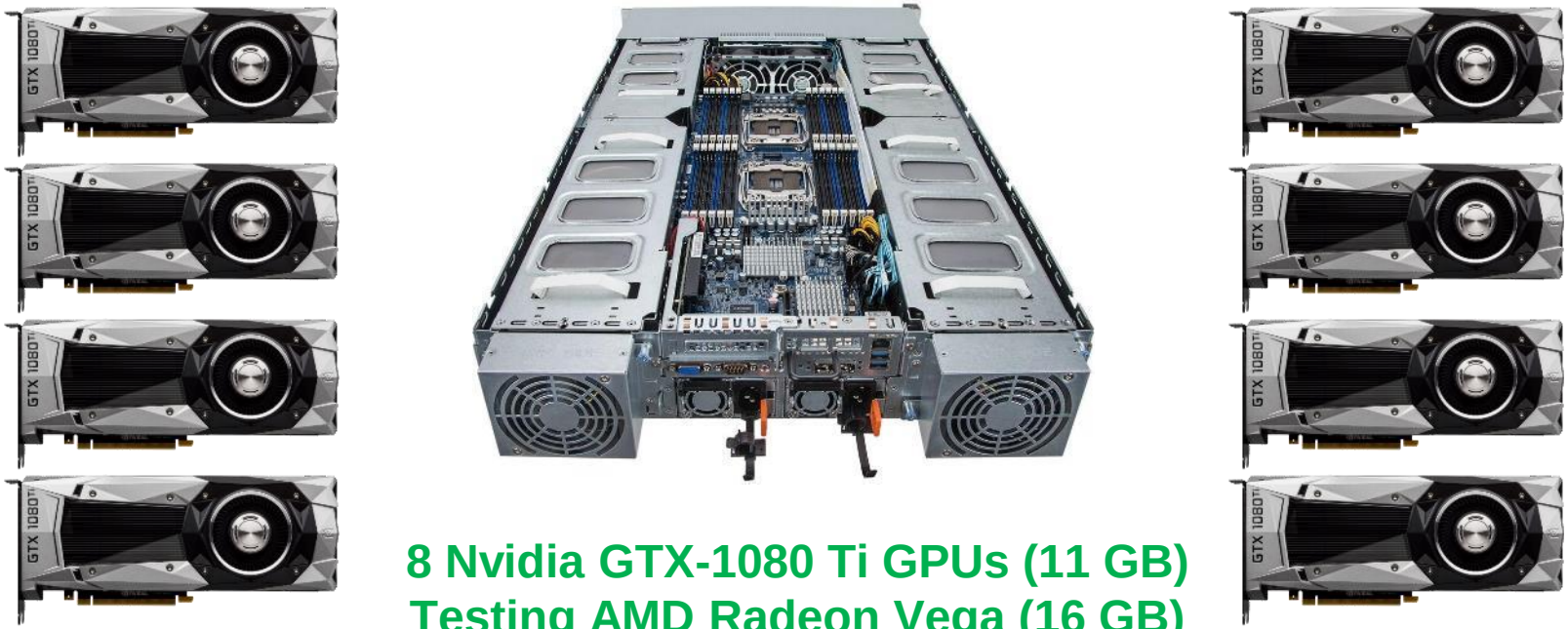
This project, called the **Cognitive Hardware And Software Ecosystem Community Infrastructure (CHASE-CI)**, will build a cloud of hundreds of affordable Graphics Processing Units (GPUs), networked together with a variety of neural network machines to facilitate development of next generation cognitive computing.

This cloud will be accessible by 30 researchers assembled from 10 universities via the NSF-funded Pacific Research Platform. These researchers will investigate a range of problems from image and video recognition, computer vision, contextual robotics to cognitive neurosciences using the cloud to be purpose-built in this project.

256 FP16/32 GPUs in 32 2U Intel Scaleable Dual 12 core with Optane Memory

Source: John Graham, Calit2

FIONA8: a FIONA with 8 GPUs
Supports PRP Data Science Machine Learning--4M GPU Core Hours/Week



8 Nvidia GTX-1080 Ti GPUs (11 GB)
Testing AMD Radeon Vega (16 GB)

24 CPU Cores, 32,000 GPU cores, 96 GB RAM, 2TB SSD, Dual 10Gbps ports
2RU; ~USD $16,000 (moving to 256GB RAM)

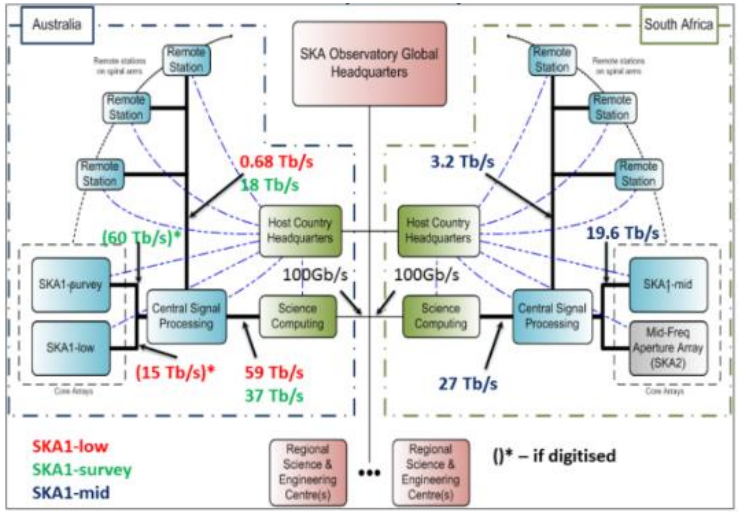
Source: John Graham, Calit2

Global Scientific Instruments Will Produce Ultralarge Datasets Continuously Requiring Dedicated Optic Fiber and Supercomputers

Square Kilometer Array

IBM to build exascale supercomputer for the world's largest, million-antennae telescope

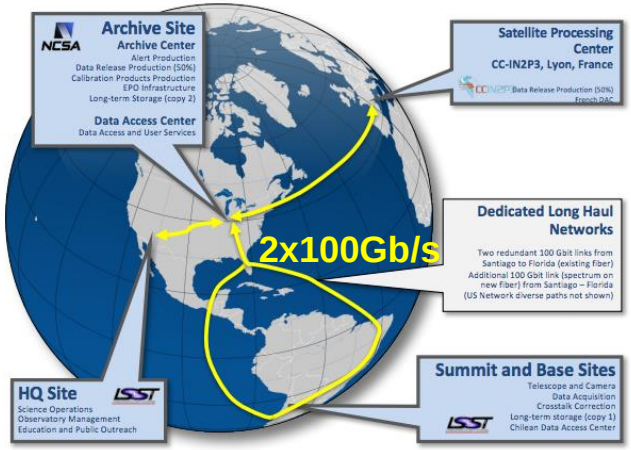
By Sebastian Anthony on April 2, 2012 at 11:48 am | 8 Comments



https://tnc15.terena.org/getfile/1939

Large Synoptic Survey Telescope

3.2 Gpixel Camera
Tracks ~40B Objects,
Creates 10M Alerts/Night
Within 1 Minute of Observing



"First Light"
In 2019

100 Gbps FIONA at UCSC Allows for Downloads to the UCSC Hyades Cluster from the LBNL NERSC Supercomputer for DESI Science Analysis

INTERMEDIATE PALOMAR TRANSIENT FACTORY

300 images per night.
100MB per raw image

120GB per night

Source: Peter Nugent, LBNL
Professor of Astronomy, UC Berkeley

Precursors to LSST and NCSA

NSF-Funded Cyberengineer
Shaw Dong @UCSC
Receiving FIONA
Feb 7, 2017

Dark Energy Spectroscopic Instrument

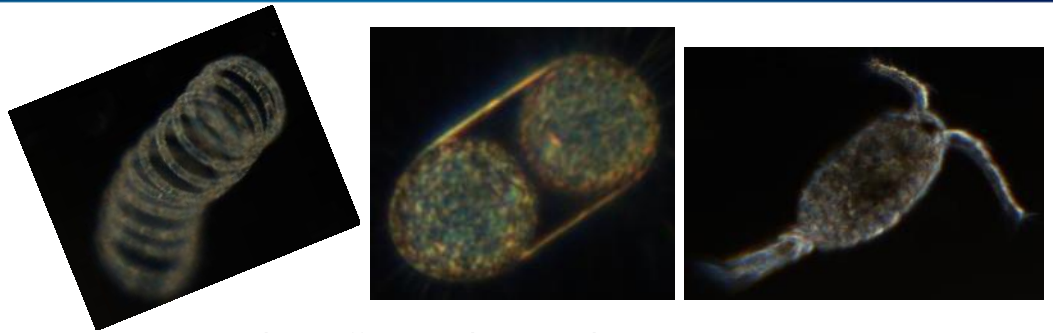
250 images per night.
530MB per raw image

800GB per night

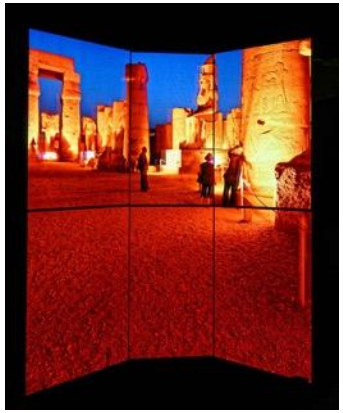
The Prototype PRP Has Attracted New Application Drivers

jupyter

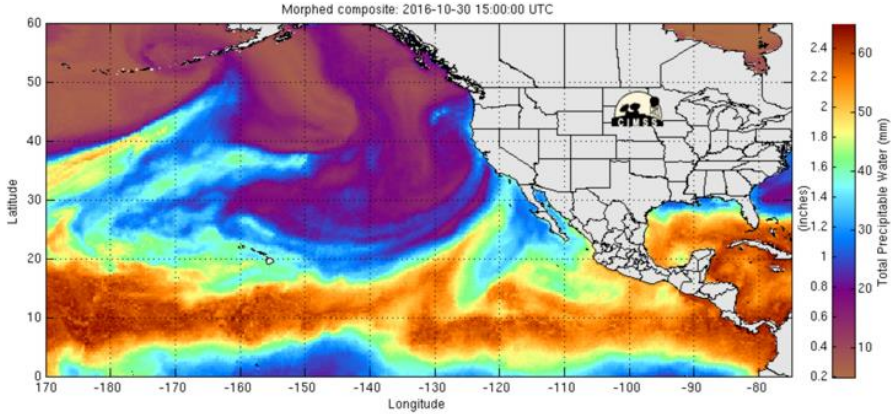
Evolved from the IPython Project



Jules Jaffe – Undersea Microscope



Tom Levy, Cultural Heritage



Scott Sellars, Marty Ralph
Center for Western Weather and Water Extremes

NSF

it²

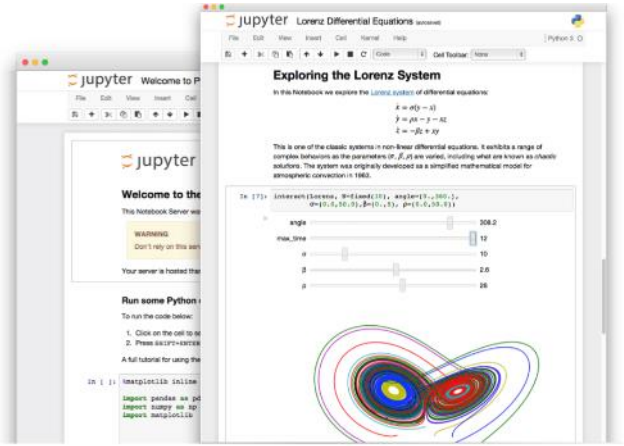
ESnet

CITRIS AND THE BANATAO INSTITUTE

PRP

CENIC

Jupyter Has Become the Digital Fabric for Data Sciences



The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Language of choice

The Notebook has support for over 40 programming languages, including Python, R, Julia, and Scala.

Share notebooks

Notebooks can be shared with others using email, Dropbox, GitHub and the Jupyter Notebook Viewer.

Interactive output

Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.

Big data integration

Leverage big data tools, such as Apache Spark, from Python, R and Scala. Explore that same data with pandas, scikit-learn, ggplot2, TensorFlow.

PRP "Broader Impacts"

- Democratizing access to data
  - Partnerships with CENIC (State of CA, Cities of Sacramento, Los Angeles)
  - Open data movements
  - Open Science Grid
- Democratizing computation
- Reaching Minority-Serving Institutions for access to data-intensive networks and training
- Collaborating with community colleges, diploma, certification and non-degree training programs

Source: Camille Crittenden, CITRIS

Challenge: CI workforce

- Retention issues particularly acute for engagement roles
- They sit somewhere between domain science and IT
- Often do not get recognition merited for this key role
- Need to think more carefully about career paths

Opportunity: Training, Workforce Development

Two teams of students at UCB, led by PRP science engagement team with partners in Research IT and ESnet.

- Team 1, working on network monitoring and visualization tools
- Team 2, working with photogrammetry, movement of large files of high-resolution video between campuses

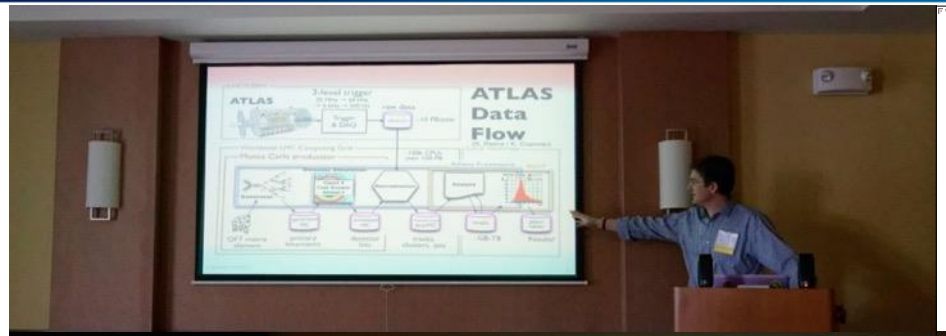
Source: Camille Crittenden, CITRIS

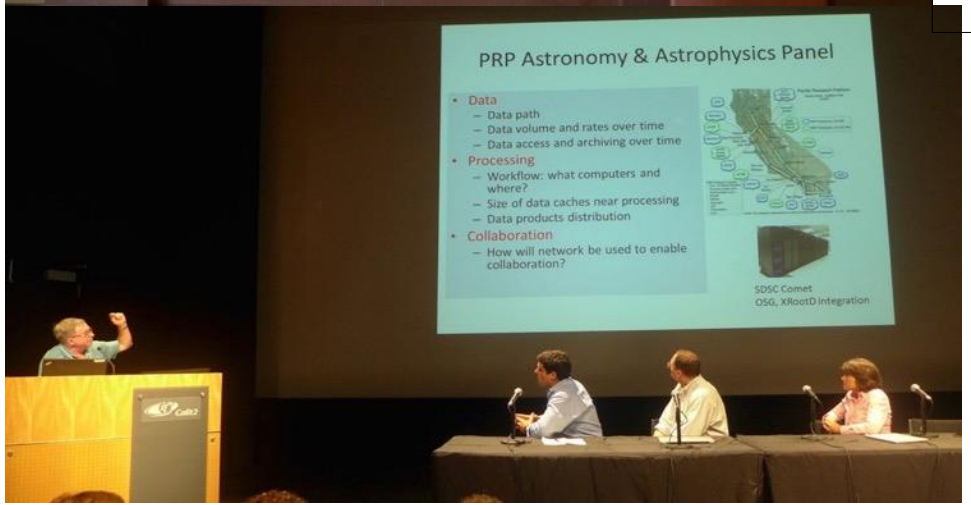
Pacific Research Platform: Future Plans for Science Engagement

- Science communications: more user-friendly engaging website (pacificresearchplatform.org). Will describe use-cases, provide updates on meetings, offer channel to contact science engagement team or network engineers.
- Students: new teams will be recruited during S18, continue work on network monitoring and digital heritage preservation projects
- Prospective workshops, domain engagement:
  - SimCenter (Center for Computational Modeling and Simulation of Natural Hazards, NSF-funded NHERI program)
  - Modeling and visualization of landslide risk (PI K. Soga)
  - Global Lives Project: 24-hour documentation worldwide
  - Electron microscopy (Nogales Lab at UCB, Villa Lab at UCSD)
- Conferences: CENIC (March), NRP v. 2 (August)

Source: Camille Crittenden, CITRIS

Pacific Research Platform: Science Engagement Workshops





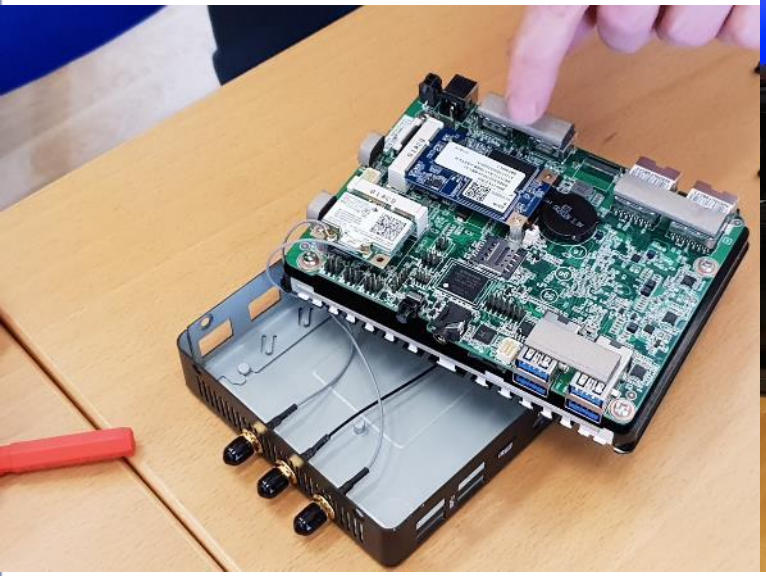
Source: Larry Smarr, Calit2

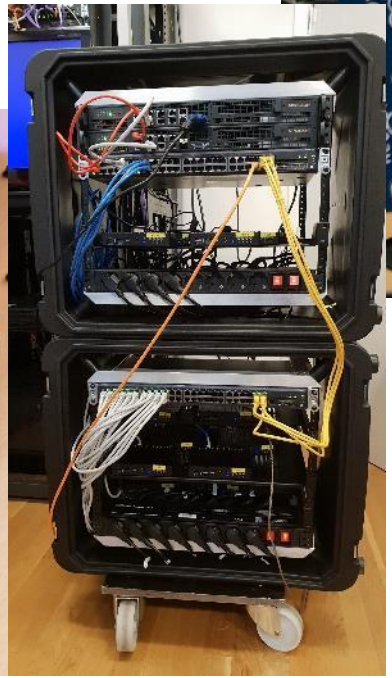
Pacific Research Platform: FIONA Workshops

Photo credit: Andrew Kirkpatrick, NSRC

Pacific Research Platform: FIONette low-cost 1Gbps DTN

Gigabyte EL-20-3700 with M.2 mSATA III SSD
iperf3 memory-to-memory > 930Mbps
GridFTP disk-to-disk > 800Mbps





Prototype flight-case:
(25) Gigabyte EL-20-3700 w/32GB SSD
(2) SuperMicro 5018 1RU 2x10G servers
(2) HPe 48x1Gbe Ethernet switch
(2) Ubiquiti UAP-AC-PRO Access Point

Photo credit: Erik-Jan Bos, NORDUnet

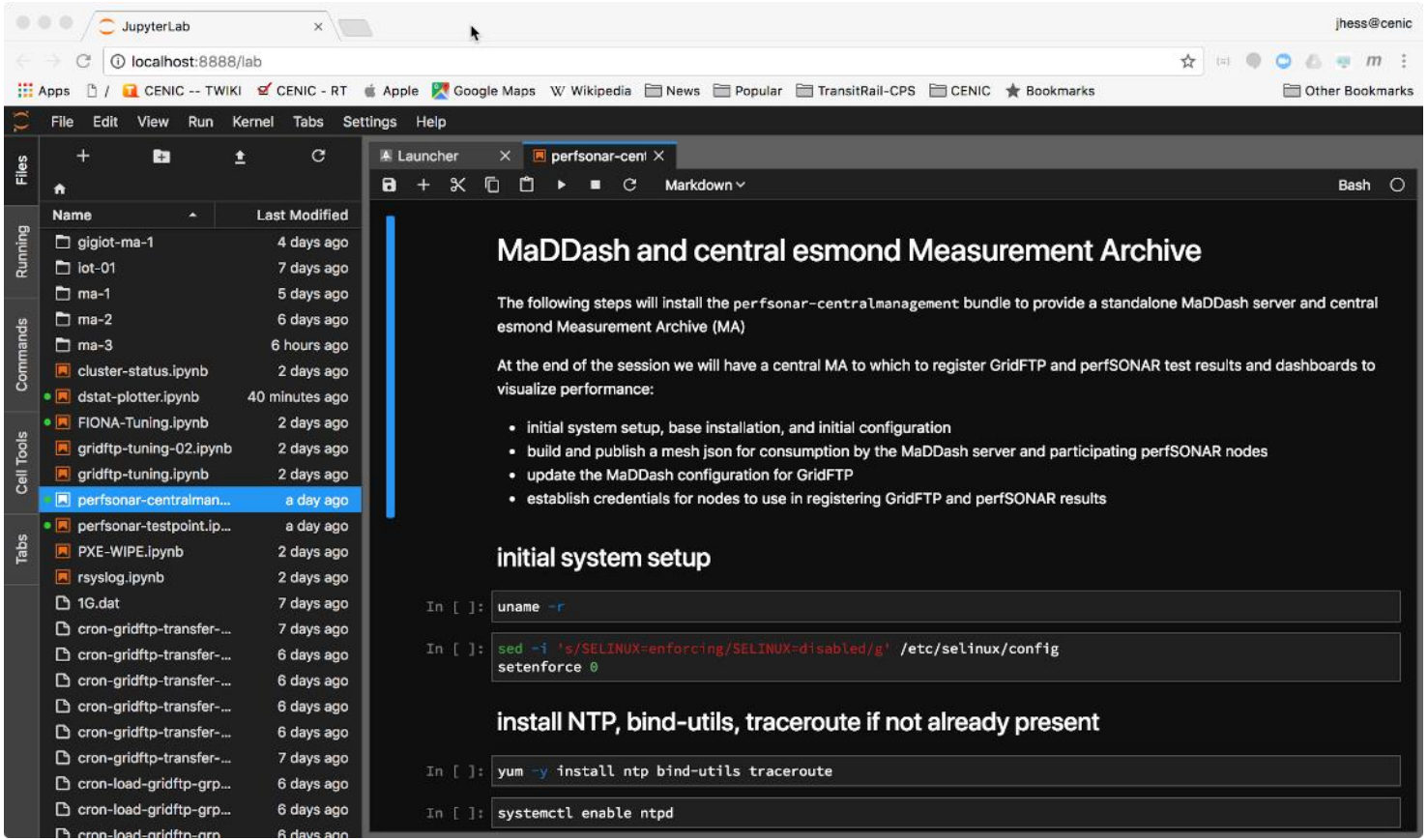
PRP FIONA Workshop: Topics and Objectives
2-5 August 2018: Bozeman, Montana

- **August 2: System Administration, Performance Measurement and Science DMZ Concepts**
  - Linux System Administration and Security Intermediate Level
  - Network Performance Measurement concepts (perfSONAR)
  - Science DMZ Architectural Model
  - Basics of Data movement (Data Transfer Nodes [DTNs] and Scientific Workflows)

- **August 3 - 4: FIONA--Advanced Performance Measurement and Visualization**
  - FIONA / FIONette setup and operation
  - Planning network performance measurement deployments (mesh; regular testing with perfSONAR tools)
  - Integrating disk-to-disk performance measurement (mesh; regular testing with GridFTP, nuttcp)
  - MaDDash and Measurement Archive (planning mesh testing regime, registering results, and visualization)
  - Troubleshooting, Tuning, and Scaling

- **August 5: Introduction to Kubernetes Container Orchestration**
  - Containers - installing docker, running containers
  - Installing single-node kubernetes (minikube), running containers in it
  - Instantiating a cluster of containers (master node, worker nodes, pod networks, pods)
  - Orchestrating persistent, distributed cluster storage (Rook / Ceph over K8s)
  - Deploying applications and performance measurement across a cluster
  - Setting up users policies, cluster security
  - Cluster, multi-campus System Administration (federated AAA; managing resources; monitoring with Prometheus, and visualizing with Grafana)

Jupyter Lab notebooks



MaDDash: combining perfSONAR and GridFTP grids on same dashboard

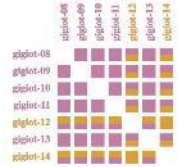
137.164.48.230/maddash-webui/

Astronomy and Astrophysics GridFTP and perfSONAR Dashboard

Astronomy and Astrophysics GridFTP - Throughput

Throughput >= 900Mbps | Throughput < 900Mbps | Throughput <= 500Mbps | Unable to retrieve data | Check has not yet run

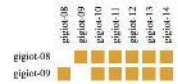
⚠ Found a total of 16 problems involving 7 hosts in the grid



Astronomy and Astrophysics perfSONAR - iperf3 Throughput, disjoint

Throughput >= 900Mbps | Throughput < 900Mbps | Throughput <= 500Mbps | Unable to retrieve data | Check has not yet run

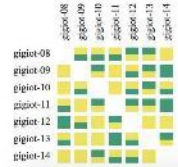
⚠ Found a total of 1 problem and it is affecting the entire grid



Astronomy and Astrophysics perfSONAR - iperf3 Throughput

Throughput >= 900Mbps | Throughput < 900Mbps | Throughput <= 500Mbps | Unable to retrieve data | Check has not yet run

✓ No problems found in grid



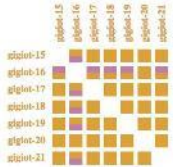
137.164.48.231/maddash-webui/

Biomedical GridFTP and perfSONAR Dashboard

Biomedical GridFTP - Throughput

Throughput >= 900Mbps | Throughput < 900Mbps | Throughput <= 500Mbps | Unable to retrieve data

⚠ Found a total of 24 problems involving 7 hosts in the grid



Biomedical perfSONAR - iperf3 Throughput, disjoint

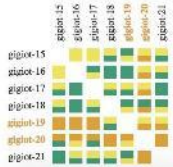
Throughput >= 900Mbps | Throughput < 900Mbps | Throughput <= 500Mbps | Unable to retrieve data

✓ No problems found in grid

Biomedical perfSONAR - iperf3 Throughput

Throughput >= 900Mbps | Throughput < 900Mbps | Throughput <= 500Mbps | Unable to retrieve data

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



GridFTP transfer metadata and data in Esmond Measurement Archive
Esmond = postgreSQL (metadata) + cassandra (data)


```
        ],
        "source": "137.164.48.207",
        "destination": "137.164.48.210",
        "measurement-agent": "137.164.48.207",
        "tool-name": "gridftp",
        "input-source": "gigiot-07.conf.cenic.org",
        "input-destination": "137.164.48.210",
        "tcp-window-size": "235104",
        "gridftp-program": "globus-gridftp-server",
        "gridftp-bytes-transferred": "1073741824",
        "gridftp-block-size": "262144",
        "bw-stripes": "1",
        "bw-parallel-streams": "1",
        "uri": "/esmond/perfsonar/archive/0d6ad42e8df94e81976c59cfc4718e81/",
        "metadata-count-total": 27,
        "metadata-previous-page": null,
        "metadata-next-page": null
    }
```

Time-series graphs: perfSONAR and GridFTP throughput on same axis

What we are working on within PRP dev

Among the technical challenges we have encountered with PRP: selectively announcing reachability of Science DMZ resources; choosing a traffic-engineered path (to make use of super-facilities, or super-channels); developing and implementing tools to ensure traffic fits within participants security model.

Within PRP dev we are exploring potential approaches:

- BGP Communities for tagging and selectively announcing routes for Science DMZ resources
- Separate ASN for PRP: Pacific Research Platform / AS395889 as a potential overlay network across R&E (similar to LCHONE VRF), as well as other approaches (see below)
- BGP pilot
  - Route Servers placed at exchange points to exchange reachability of resources
  - IPv6 only (may support IPv4 as transport)
  - Initial phase will include UCSD, SDSC, UCSC, Stanford, NCSA, UIC/EVL
- BGP + SDN/SDX pilot exploring dynamically provisioned 'super-channels' supporting data movement among cooperating research groups
- Deployment of Tstat on PRP-managed DTNs
- AutoGOLE / NSI running within containers across Nautilus cluster
- Kubernetes namespaces for per-group federated authorization, access to distributed cluster resources

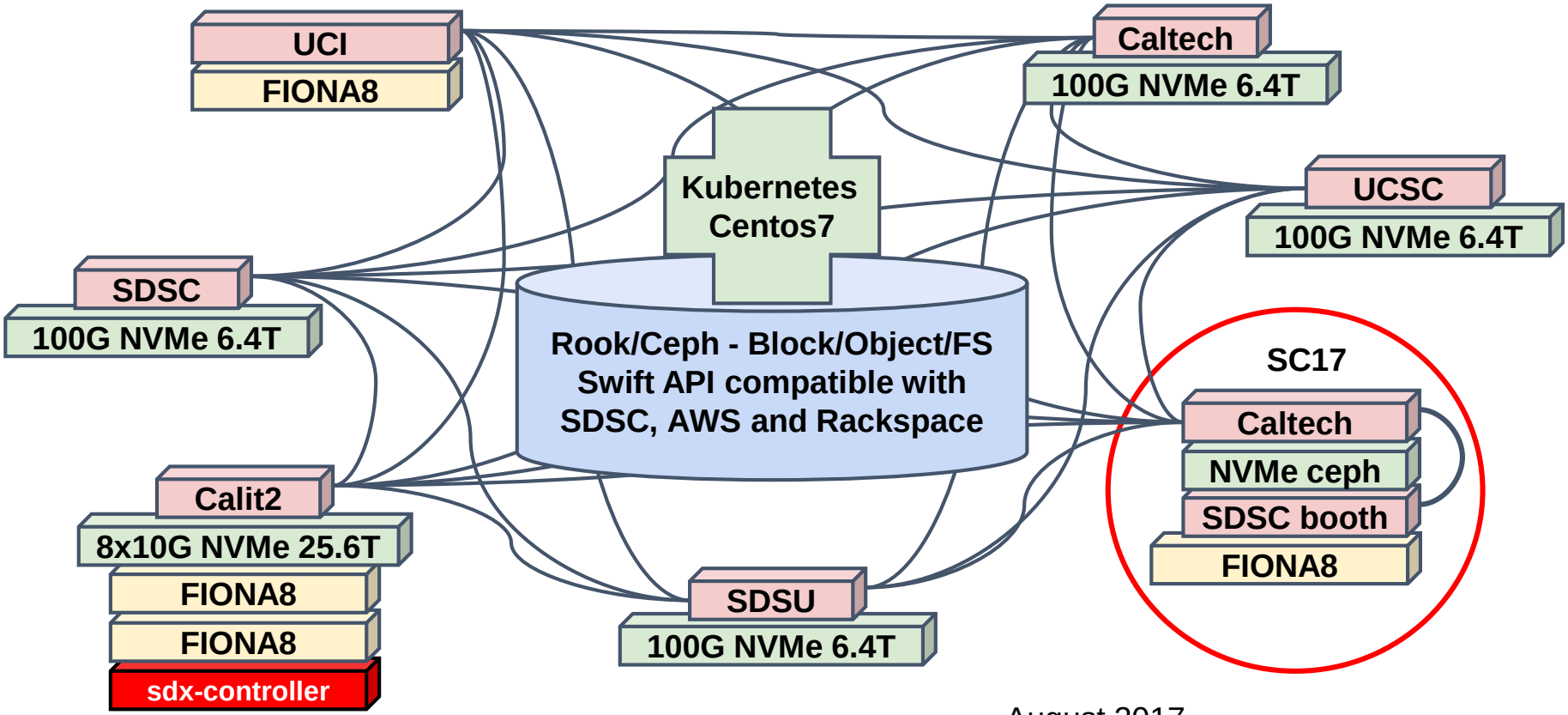
Multi-Institution Hyper-Converged ScienceDMZ

John Graham
Research and Development Engineer
Calit2/Qualcomm Institute, UCSD
jjgraham@eng.ucsd.edu

Dmitry Mishin, PhD.
Applications Programmer
SDSC, UCSD
dmishin@ucsd.edu

https://docs.google.com/presentation/d/11WblUmRhh5O9rpiWfQ5gbq2ewmLZAFKeBurkwUsp2rw/edit?usp=sharing

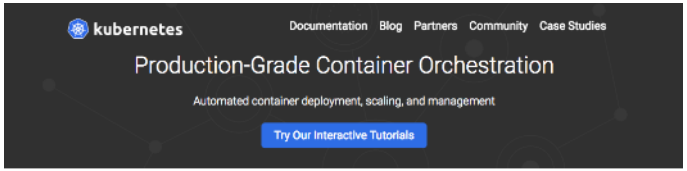
Multi-Institution, Hyper-Converged, Cloud Native Storage



August 2017

Source: John Graham, Calit2

Rook is Cloud Native Ceph in our Hyper-converged cluster



Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications.

It groups containers that make up an application into logical units for easy management and discovery. Kubernetes builds upon 15 years of experience of running production workloads at Google, combined with best-of-breed ideas and practices from the community.

Planet Scale

Designed on the same principles that allows Google to run billions of containers a week, Kubernetes can scale without your ops team.

Never Outgrow

Whether testing locally or running a global enterprise, Kubernetes flexibility grows with you to deliver your applications consistently and easily no matter how complex your need is.

Run Anywhere

Kubernetes is open source giving you the freedom to take advantage of on-premise, hybrid, or public cloud infrastructure, letting you effortlessly move workloads to where it matters to you.

GPU

+

NVMe

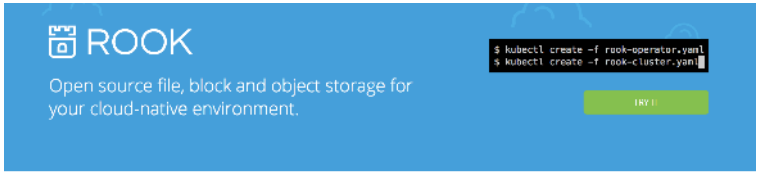
+

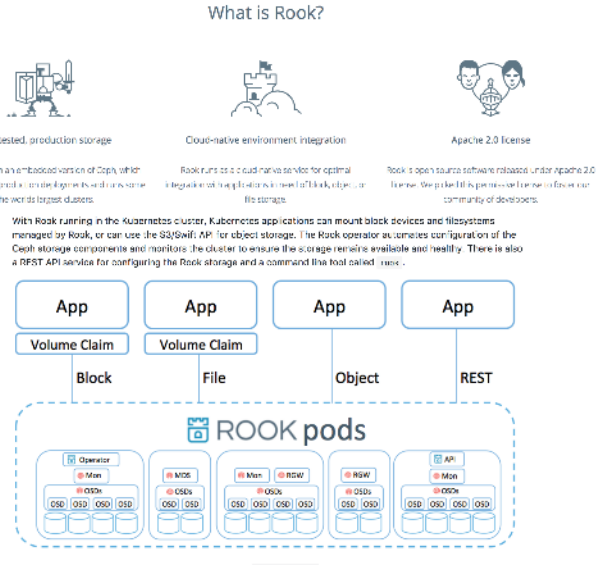
40G or 100 NIC

=

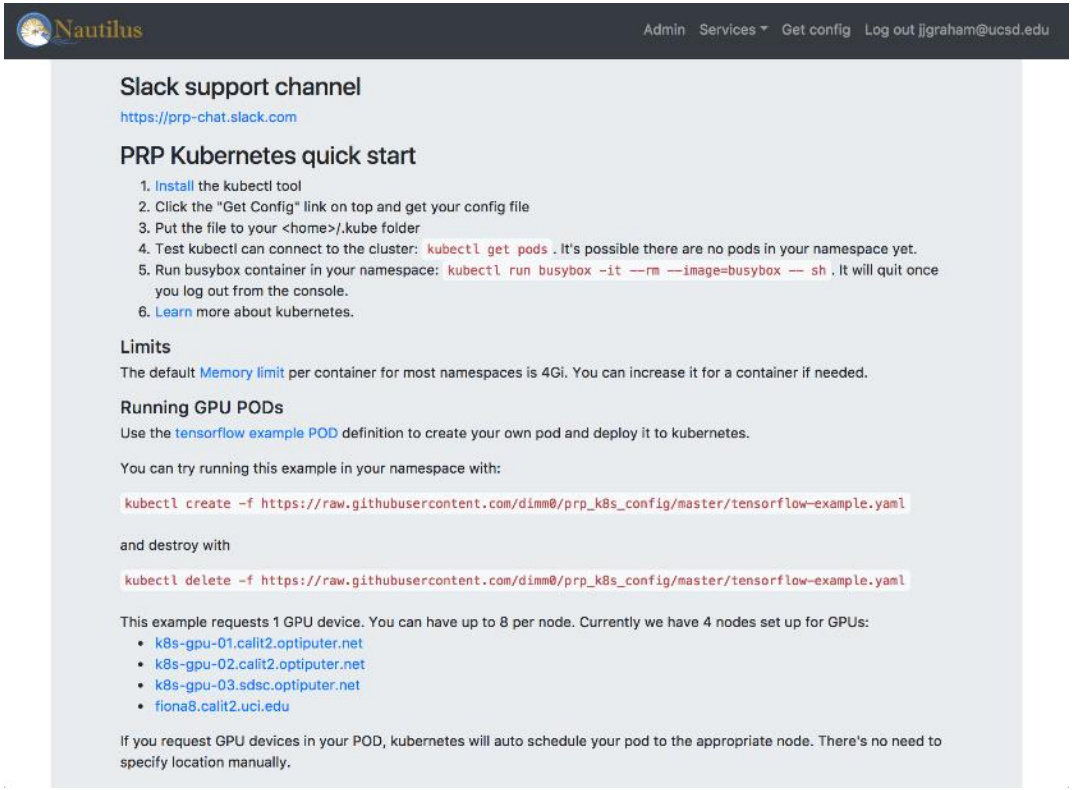
FIONA8

Source: John Graham, Calit2





Nautilus Hypercluster Portal



- CIlogon OIDC authentication
- Grafana and Prometheus
- MaDDash automation
- Traceroute Visualization
- Admin interface
- kubectl config download
- Node and Pod dashboards
- InMon services
- ...

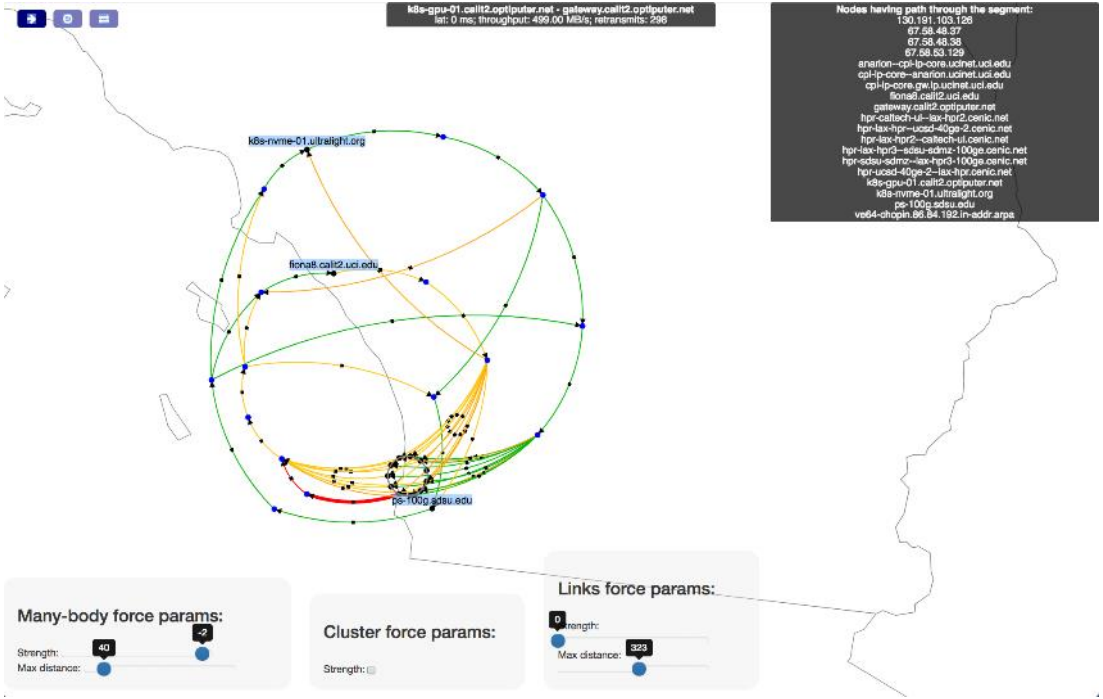
Source: John Graham, Calit2

Grafana Dashboard for the UChicago ATLAS GAN simulations

Source: John Graham, Calit2

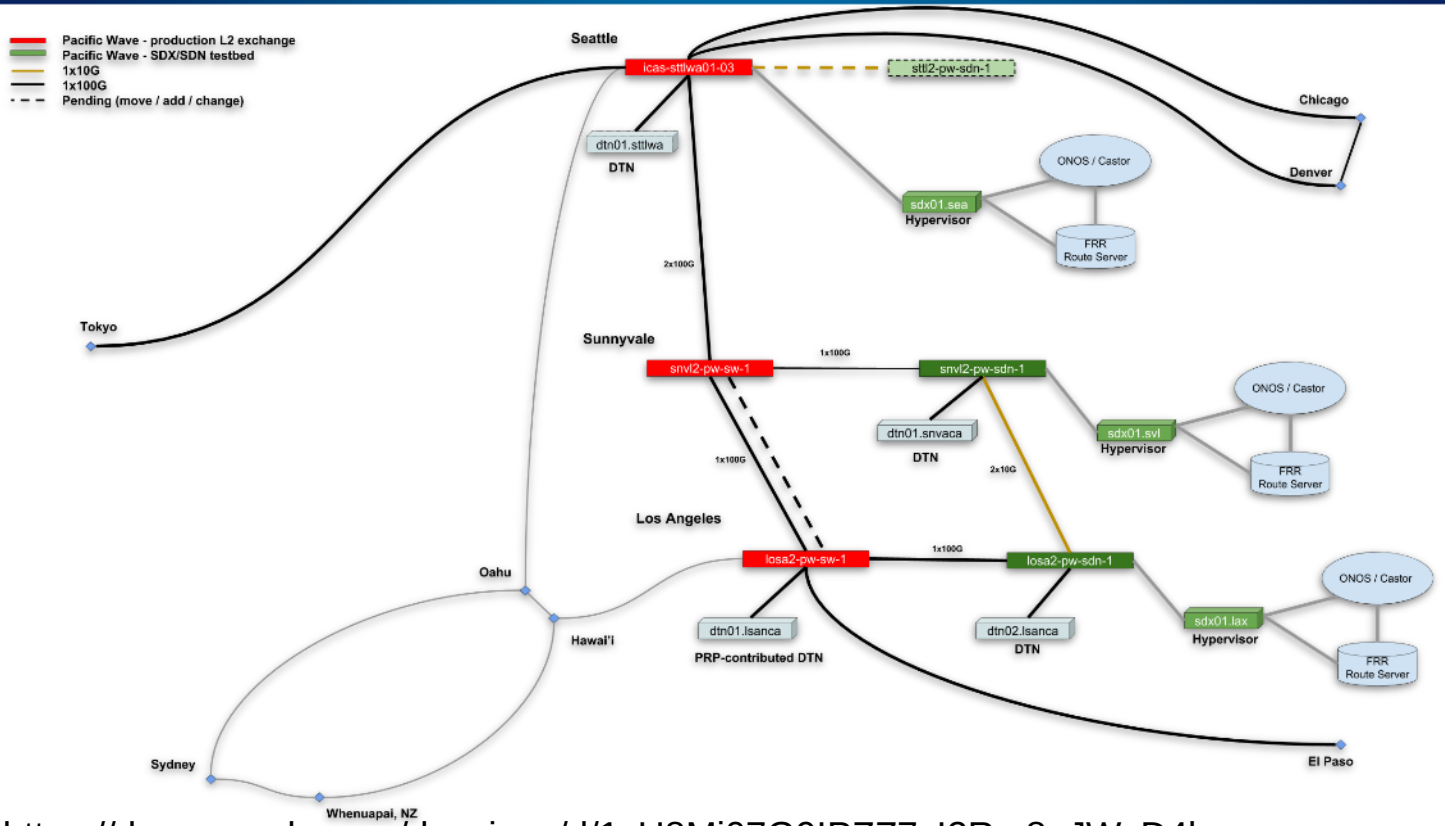
MaDDash and Traceroute visualization tools

https://perfsonar.nautilus.optiputer.net/maddash-webui/



https://traceroute.nautilus.optiputer.net/

Pacific Wave: IRNC SDX/SDN testbed control plane

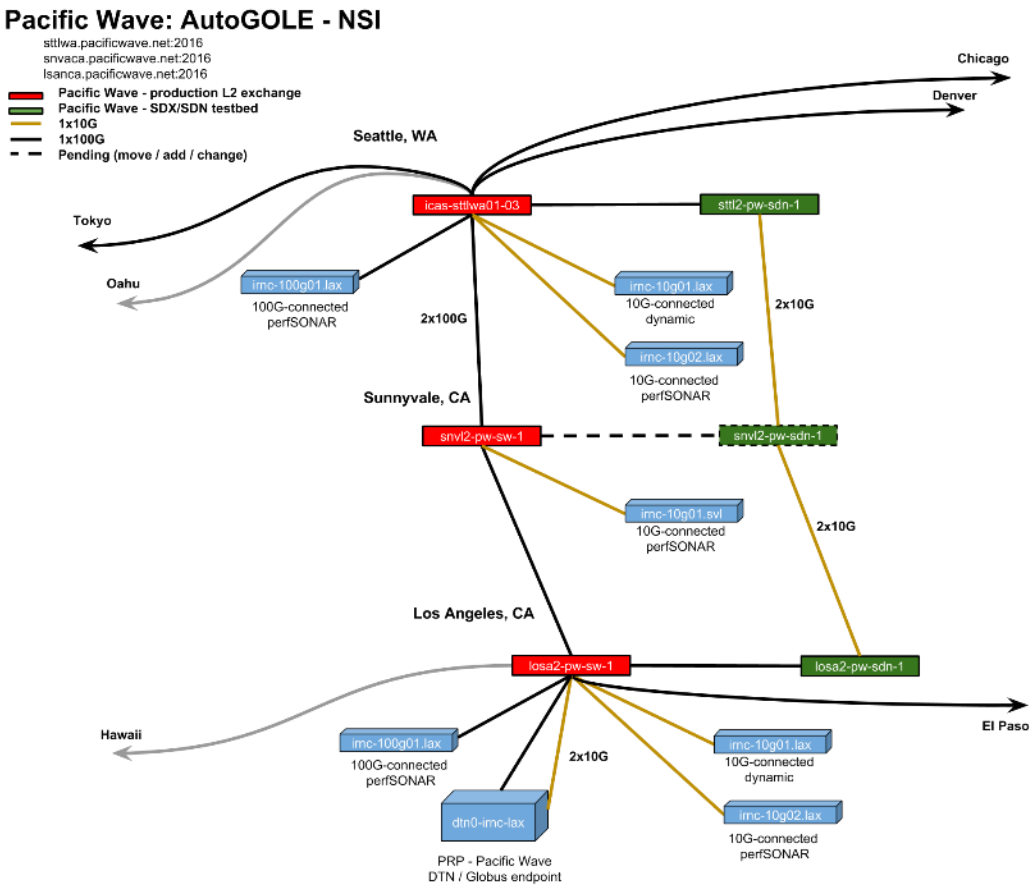


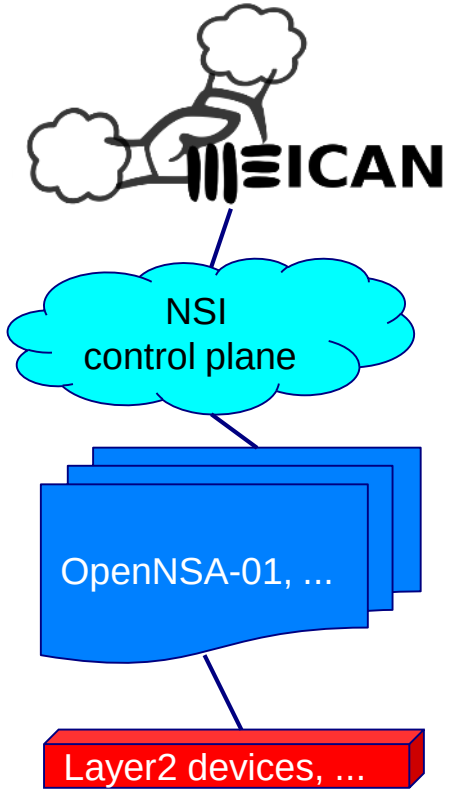
https://docs.google.com/drawings/d/1xU8Mi07G9IPZZ7vI2Rm8oJWsD4ln-D_slN8rbFAwwc4/edit?usp=sharing

AutoGOLE Fabric

https://drive.google.com/open?id=1TubFfvM2iT80tF67P7AZHnx5T8O736CO&usp=sharing

Pacific Wave participation in AutoGOLE / NSI + MEICAN pilot





Scaling the PRP -> NRP, xRP, GRP

The 2nd National Research Platform Workshop will focus on how local, state, regional and national groups can effectively collaborate to scale PRP and realize the vision of a true National Research Platform:

http://www.cvent.com/events/national-research-platform-conference-toward-a-national-big-data-superhighway/event-summary-48a69b9807bd46ecb5d4343bcbfa61c5.aspx

2NRP Program overview

- Progress and success stories since last year's NRP workshop
- Building and scaling a common hardware platform and a common software stack
- Highlighting applications and scientific use cases and attracting new users
  - Solutions and challenges in supporting scientific collaborations as a community
  - Scaling infrastructure for trusted collaboration for distributed Virtual Organizations (VOs), research groups and end users
- Enabling access and usability by small institutions through outreach, such as research facilitators and training
- Scaling the NRP:
  - Network architecture
  - Monitoring and measurement
  - Security
  - Near and long-term potential NRP capabilities
- Scaling and coordination of existing local, state, regional and national infrastructure

CI that benefits from integration

- Need **simple** apps for data mobility
- Requires **partnerships** between researchers and CI experts
- **Disconnected** collaborators, often in separate institutions, separate funding, and may not interact frequently
- The ScienceDMZ/DTN **architecture** is an effective means of enabling high-performance end-to-end networking for campuses and institutions, balancing researcher requirements and network security concerns. Many DTNs, and FIONA models are cost-effective solutions.
- The Science Engagement process is crucial to **scaling** up to a national/international research platform, to identify real requirements, design and build to match

Source: Camille Crittenden, CITRIS

Indicators for Success

- Voluntary institutional and researcher commitment is vital
- Corollary: better to be bottom-up rather than top-down
- Scaling will be hard: end-to-end perf. work can be $O(N^2)$

Tools and protocols needed for scaling up. Some ideas:

- Leverage existing orgs (ACI-REF, CaRC, XSEDE, PEARC).
- Work towards NRP by replicating PRP at regional level
- Lower barriers and efforts to adopt NRP ideas. E.g., develop minimal standard infrastructure, configurations, deployment and testing

Source: Camille Crittenden, CITRIS

Needed: Trust, collaboration

It took time for PRP participants to work together

- to learn individual roles and strengths (and weaknesses),
- to learn to rely on/trust their collaborators

Trust is a human-intensive endeavor, one relationship at a time, not readily amenable to scaling. But can foster:

- Identify and document successful collaborations (like PRP)
- Emphasize peer to peer communications (at all levels)

Source: Camille Crittenden, CITRIS

Questions? Interested in learning more? Shall we collaborate?

I would be happy to take questions -- I may even have answers.

If you heard something which interests you about the Pacific Research Platform and would like to learn more, please visit:

http://pacificresearchplatform.org

PRP Engineering Calls, Thursdays, 10:00-11:00 Pacific: https://cenic.zoom.us/j/997313186

PRP-L listserv: https://mailman.ucsd.edu/mailman/listinfo/prp-l

PRP Slack Channel: https://prp-chat.slack.com

Nautlius Rocket.chat: https://rocket.nautilus.optiputer.net

Are working on a similar or related project and have something to share, let's collaborate!

19º WRNP
Workshop RNP
7 | 8 MAIO
Campos do Jordão | SP

Obrigado

John Hess
jhess@cenic.org