

Machine Learning Research Facilitated by UCSD's Research IT Services Team: Supporting Affordable Cloud Storage/ GPU Computing

Tom DeFanti

Research Scientist, Co-PI *The Pacific Research Platform, and CHASE-CI*

Project Manager of *Towards the National Research Platform*

Qualcomm Institute/California Institute for Telecommunications and Information

University of California San Diego

Distinguished Professor Emeritus, University of Illinois at Chicago

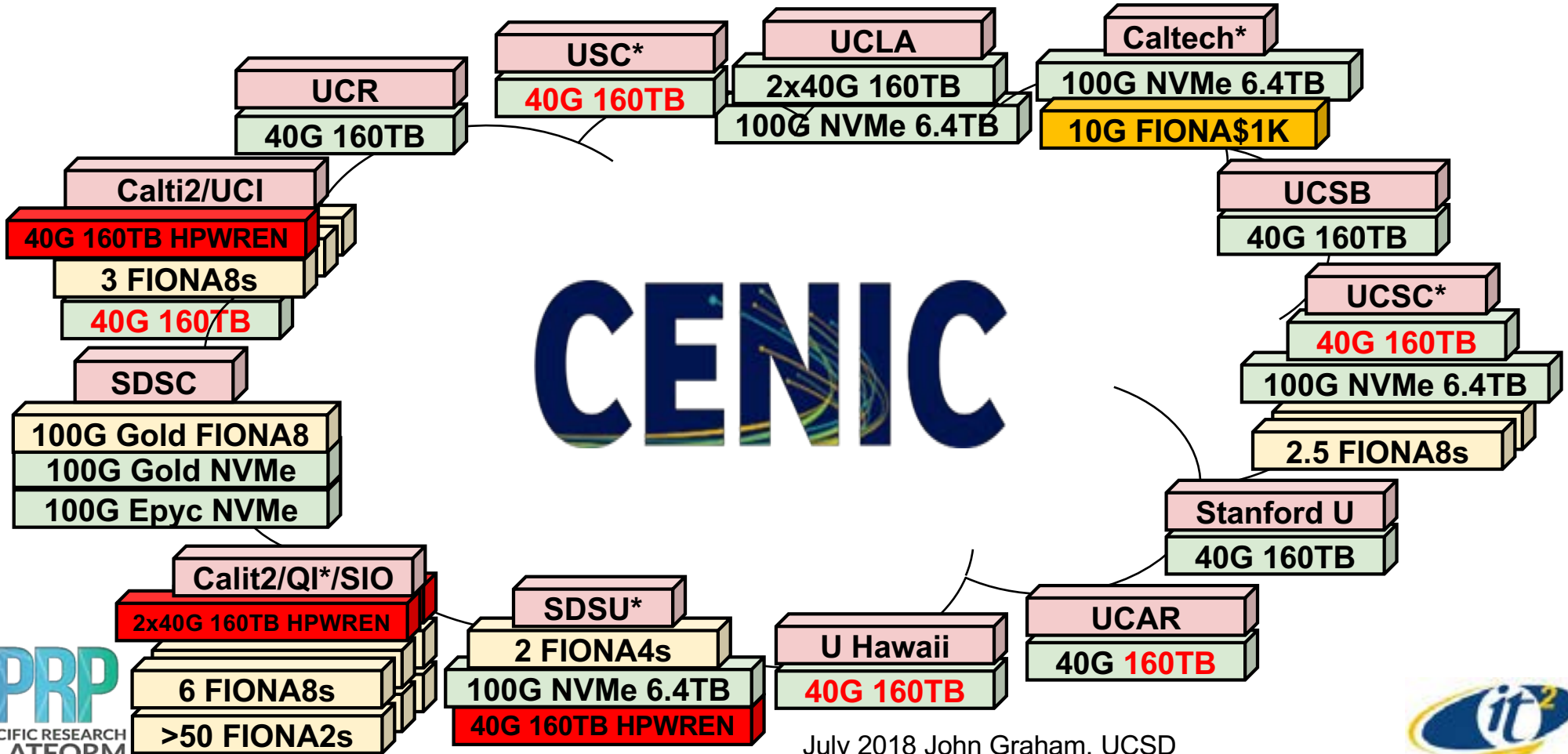


Nautilus: A High-Performance, Low-Cost California Community Cluster

- Our research has developed purpose-built 'FIONA' rack-mounted PCs that are tuned to test end-to-end 1G, 10G, 40G and 100G connections.
- We are using Kubernetes for testing and visualizations as well for managing community-based CPU/GPU/Storage called *Nautilus*, a distributed cluster of FIONAs
- Our Research IT Services team helps us by supporting the campus PRISM research network and its connection to CENIC's High Performance Research Network.
- This has helped us get 4 major grants (over \$10M for CI development)
- UC San Diego's Research IT Services Team clones our working tech.

Nautilus October 2018: Storage and GPUs

CENIC



Installing 16 10&12 TB Drives in June at UC Merced, UC Riverside, and Stanford—160TB to 192TB per FIONA



Research and Instructional IT Services use Kubernetes to Manage FIONAs

“Kubernetes is a way of stitching together a collection of machines into, basically, a big computer,”

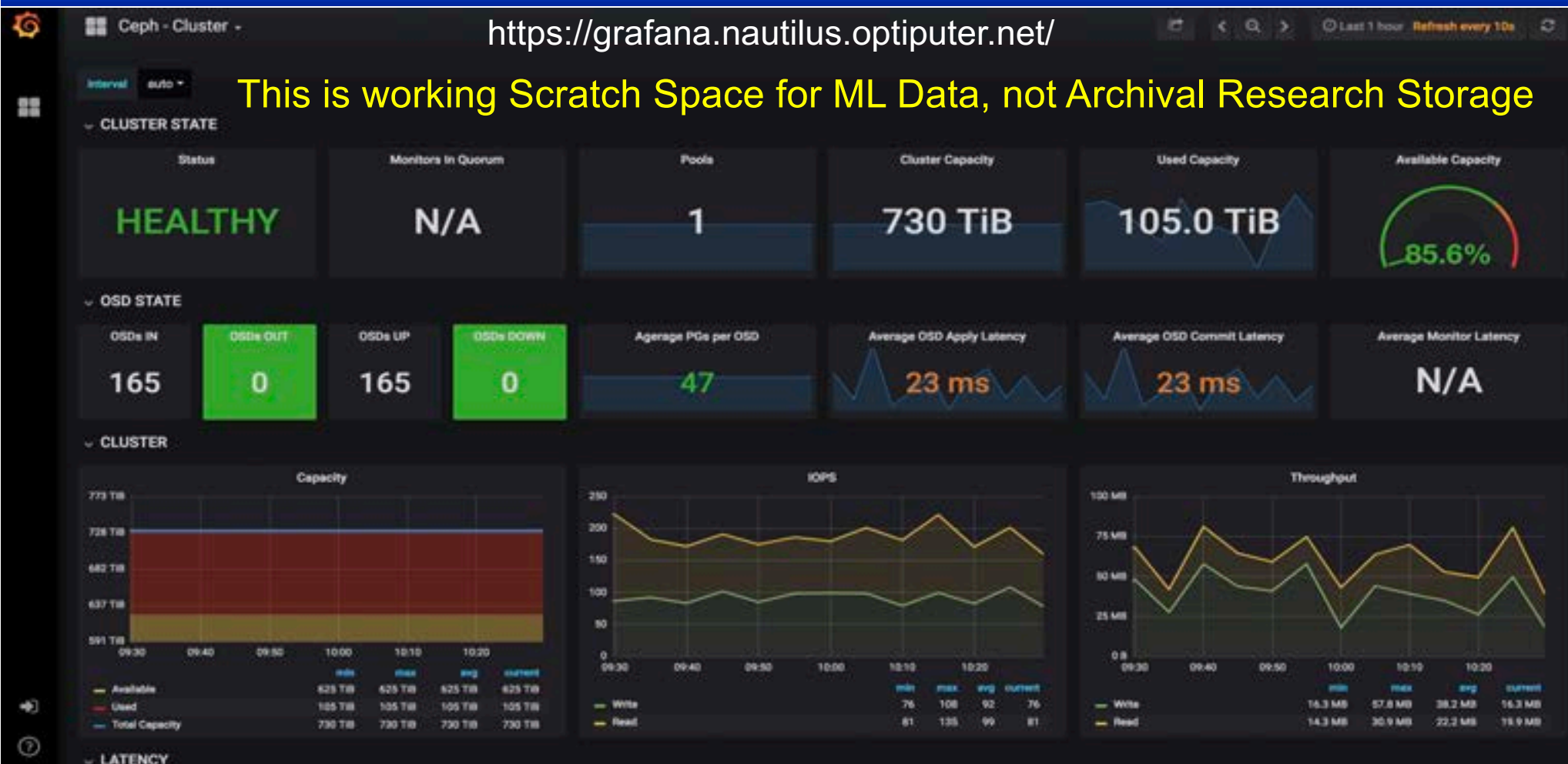
--Craig Mcluckie, Google
and now CEO and Founder of Heptio

Kubernetes with Rook/Ceph allows us to Manage Petabytes of Distributed Storage and GPUs for Data Science as well as Measure and Monitor Network Use

Grafana Plot of First 730 TB at UCSD, Stanford, UCD, UCM

<https://grafana.nautilus.optiputer.net/>

This is working Scratch Space for ML Data, not Archival Research Storage



ML GPU Users: at least 15-1 Price Advantage at UC San Diego vs. AWS Our Research IT Services Team Provides Space and Power

| Nvidia Card | ~Cost | 32-bit GF | GB | per GF | per GB | cores | 8-GPU PC |
|---|----------|-----------|----|--------|--------|-------|-----------|
| GTX 1080-Ti 11GB | \$789 | 10609 | 11 | \$0.07 | \$72 | 3584 | \$20,311 |
| V100 16GB | \$11,418 | 14899 | 16 | \$0.77 | \$714 | 5120 | \$105,344 |
| AWS p2.xlarge EC2 (8) K-80 GPUs+disk for 4 years | | | | | | | \$318,720 |
| AWS p2.xlarge EC2 (8) K-80 GPUs+disk for 4 years +57% ICR | | | | | | | \$500,390 |



**40 GPUs in 5
FIONA8s for ML in
15" of rack space
2.4 Million GPU
hours per day**

Grafana Plot of First 93 GPUs Online



Instructional Data Science/Machine Learning Platform

IT SERVICES

UCSD *Cognitive Science*

UC San Diego
Jacobs School of Engineering

Student-focused GPU cluster supporting:

- Coursework across all disciplines
- For-credit student research

Aligned with Research best practices

Managed by UCSD IT Services:

- Academic Technology Services
 - Research IT Facilitation & Integration Services
 - Educational Technology Services
- Educational Computing Environments

- *Leverage existing instructional IT staff, skill-sets & technologies*
- *Create a low-friction transition of tools from lab to classroom and back*

Instructional Data Science/Machine Learning Platform

Hardware/OS Configuration



- Ten FIONA8 Nodes:
 - 24-32 cores
 - 256-384 GB RAM
 - 8x GTX 1080Ti GPU
- Two Legacy GPU Nodes:
 - 15x Titan-Kepler
- 20TB Flash Storage
- 10G networking
- Housed in SDSC COLO
- Kubernetes + Docker runtime
- CentOS7 + Puppet
- ZFS/NFS Filesystem

- *Students need 10's of GB storage, not TBs*
- *Increase RAM - allow more students debug/create code w/o GPU*
- *CentOS7 to match our instructional Linux env.*
- *Accept reduced performance for simplified user/sysadmin experience (e.g. POSIX NFS, not Block; 10G vs 40/100G net)*

More FIONA8s for Research Coming

- UC Berkeley estimates they spend \$40,000 in cloud GPU per published grad student paper (!). Same as 2 FIONA8s with 256GB & 32 cores each
- ECE Department has allocated funds to buy 4 FIONA8s this year, 4 more next year—that's 64 GPUs coming for ECE faculty research
- CSE faculty are proposing to buy FIONA8s to add to Nautilus; Prof. Vineet Bafna already has
- Our CHASE-CI grant has funding for many more FIONA8s for faculty
- We need the Research IT Services team to rack & network & sustain these in the SDSC colo space like Instructional IT services does.
- Other campuses are building similar/identical systems for their data sciences institutes—the competition for research funding is real

We Thank Our Supporters:

- **UC San Diego IT Services:**
 - Research IT Services
 - Ed Tech Services
 - The Next Generation Network program
- **University of California Office of the President CIO**
- **US National Science Foundation (NSF) awards:**
CNS-1730158, ACI-1540112, ACI-1541349, & OAC-1826967
- **CENIC, Pacific NorthWest Gigapop (PNWGP), Pacific Wave & StarLight**
- **Calit2 and Calit2's Qualcomm Institute**

