Next Generation Integrated Architecture
SDN Ecosystem for LHC and Exascale Science

LHC Beyond the Higgs Boson
LSST  SKA
BioInformatics
Earth Observation
Gateways to a New Era

Harvey Newman, Caltech
NSF CC*/CICI Workshop: Data Integration Panel
Albuquerque, NM October 3, 2017
A New Era of Challenges and Opportunity

Meeting the Needs for LHC Run2 and Beyond

Paving the Way for Other Science Programs
A New Era of Challenges: Global Exabyte Data Distribution, Processing, Access and Analysis

- **Exascale Data**
  - 0.5 EB now, 1 EB by end of LHC Run2
  - To ~100 EB during HL LHC Era
- **Network Total Flow of >1 EB this Year**
  - 850 Pbytes flowed over WLCG in 2016
- **Projected Shortfalls by HL LHC**
  - CPU ~4-12X, Storage ~3X, *Networks*
- **Network Dilemma:**
  Per technology generation (~8 years)
  - Capacity at same unit cost: 4X
  - Bandwidth growth: 30X (Internet2); 50X (GEANT), 70X (ESnet)
  - *During LHC Run3 (~2022) we will likely reach a network limit*
- This is unlike the past
  - Optical, switch advances are evolutionary; *physics barriers ahead*

**New Levels of Challenge**

- **Global data distribution, processing, access and analysis**
- **Coordinated use of massive but still limited *diverse* compute, storage and network resources**
- **Coordinated operation and collaboration *within and among* scientific enterprises**
- **HEP will experience increasing Competition from other data intensive programs**
  - Sky Surveys: LSST, SKA
  - Next Gen Light Sources
  - Earth Observation
  - Genomics; Cryo-EM
Opportunity: Exploit the Synergy among

1. Global operations data and workflow management systems developed by HEP programs, *to respond to both steady state and peak demands*
   - Evolving to work with *increasingly diverse (HPC) and elastic (Cloud) resources*

2. Deeply programmable, agile software-defined networks (SDN), emerging as multidomain network operating systems (e.g. SENSE & SDN NGenIA; Multidomain multicontroller SDN)

3. Machine Learning, modeling and game theory:
   - Extract key variables; optimize; move to real-time self-optimizing workflows with Reinforcement Learning.
Quad Chart: **SENSE - SDN for End-to-end Networked Science at the Exascale**

**CHALLENGES:** Distributed scientific workflows need end-to-end automation so the focus can be on science; not infrastructure
- Manual provisioning and infrastructure debugging takes time
- No service consistency and No service visibility or automated troubleshooting across multiple domains
- Lack of realtime information from domains impedes development of intelligent services

**SOLUTIONS**
- **Model Driven SDN Control with Orchestration**
- **Intent based science application** facing APIs with resource discovery and negotiation
- Automated end-to-end troubleshooting and debugging
- **Datafication** of cyberinfrastructure to enable intelligent services

**SCIENTIFIC IMPACT:** Enables distributed science workflows to effectively orchestrate network, compute, and storage resources in real-time
- Workflow optimization based on end-to-end resource discovery and performance information
- Decision Making based on real-time interaction and negotiation
- Intent-based models & APIs enable science application workflows to interact intuitively with networks as a 1st class resource

**TEAM**
- Esnet/NERSC/LBL
- Caltech
- Fermi National Accelerator Laboratory
- Argonne National Laboratory
- University of Maryland/Mid-Atlantic Crossroads
CC* Integration: SANDIE: SDN Assisted NDN for Data Intensive Experiments

SOLUTIONS + Deliverables
- Deploy NDN edge caches with SSDs & 40G/100G network interfaces at 7 sites; combine with larger core caches
- Simultaneously optimize caching (“hot” datasets), forwarding, and congestion control in both the network core and site edges
- Development of naming scheme and attributes for fast access and efficient communication in HEP and other fields

SCIENTIFIC and BROADER IMPACT
- Lay groundwork for an NDN-based data distribution and access system for data-intensive science fields
- Benefit user community through lowered costs, faster data access and standardized naming structures
- Engage next generation of scientists in emerging concepts of future Internet architectures for data intensive applications
- Advance, extend and test the NDN paradigm to encompass the most data intensive research applications of global extent

TEAM
- Northeastern
- Caltech
- Colorado State
- In partnership with other LHC sites and the NDN project team
Responding to the Challenges

New Overarching “Consistent Operations” Paradigm

- VO Workflow Orchestration systems that are real-time and reactive ✓;
  - New: Making them *Deeply network aware, and proactive*

- *Network Orchestrators with similar, real-time character*

- Together responding moment-to-moment to:
  - State changes in the networks and end systems
  - Actual versus estimated transfer progress, access IO speed

- Prerequisites:
  - End systems, data transfer applications and access methods capable of high of throughput [Fast Data Transfer (FDT)]
  - Realtime end-to-end monitoring systems [End sites + networks]

- Modern Elements for efficient operations within the limits:
  - SDN-driven bandwidth allocation, load balancing; caching; NDN;
    - flow controls at the network edges and in the core

- Result: Best use of available network, compute and storage resources while avoiding saturation and blocking other network traffic

- Machine Learning for Optimization; Game Theory for Stable Solutions
Next Generation Integrated Architectures For Science

A New Era of Challenges and Opportunity

Meeting the Needs for LHC Run2 and Beyond

Backup Slides Follow
DOE Program: ASCR SDN-Enabled Terabits Optical Networks for Extreme-Scale Science

Program Area: SDN-enabled High Performance Networks

PI: Inder Monga (ESnet)  Deputy PI: Chin Guok
co-PIs: Harvey Newman (Caltech), Phil DeMar (FNAL), Linda Winkler (ANL), Tom Lehman (UMD/Max)

Project Title: SENSE: SDN for End-to-end Networked Science at the Exascale

Harvey Newman
Caltech
Professor of Physics
newman@hep.caltech.edu

Phil DeMar
Fermilab
Network Analysis and Architect Manager for the Core Computing Division
demar@fnal.gov

Linda Winkler
Senior Network Engineer Mathematics and Computer Science (MCS)
lwinkler@mcs.anl.gov

Tom Lehman
Director of Research
University of Maryland
Mid-Atlantic Crossroads
tlehman@umd.edu

Chin Guok
Senior Network Engineer
ESnet
chin@es.net
SENSE: SDN for End-to-end Networked Science at the Exascale
ESnet Caltech Fermilab LBNL/NERSC Argonne Maryland

- **Mission Goals:**
  - Improve end-to-end performance of science workflows
  - Enabling new paradigms: e.g. creating dynamic distributed ‘Superfacilities’.

- **Comprehensive Approach:**
  - An end-to-end SDN Operating System, with:
    - Intent-based interfaces, providing intuitive access to intelligent SDN services
    - Policy-guided E2E orchestration of resources
    - Auto-provisioning of network devices and Data Transfer Nodes
    - Network measurement, analytics and feedback to build resilience
Program, Area: CC, Integration    Award Number: 1659403
PI: Edmund Yeh
co-PIs: Harvey Newman, Christos Papadopoulos
Project Title: SANDIE: SDN-Assisted NDN for Data Intensive Experiments
HEPCloud Facility: Fermilab Bursting to the Cloud. Doubling CMS Peak Compute Capacity

Meeting the Challenge of Bringing Diverse Resources to Bear

Cores from Google

O. Gutsche
HSF Workshop
January 2017

Issue: Cloud Business Model for Data + Network Intensive Use Cases
LSST + SKA Data Movement
Upcoming Real-time Challenges for Astronomy

3.2 Gigapixel Camera (10 Bytes / pixel)

- **Planned Networks:** Dedicated 100Gs (Spectrum !) for image data, +100G for other traffic, and ~40G and up for diverse paths
- **Lossless compressed Image size = 2.7GB**
  (~5 images transferred in parallel over a 100 Gbps link)
- Custom transfer protocols for images (UDP Based)
- **Real-time Challenge:** delivery in seconds to catch cosmic “events”
- **+ SKA in Future:** 3000 Antennae covering > 1 Million km2; 15,000 Terabits/sec to the correlators ➔ 1.5 Exabytes/yr Stored
A New “Consistent Operations” Paradigm for HEP and Exascale Science

**METHOD:** Construct autonomous network-resident services that dynamically interact with site-resident services, and with the experiments’ principal data distribution and management tools.

**To coordinate use of network, storage + compute resources, using:**

1. Smart middleware to interface to SDN-orchestrated data flows over network paths with allocated bandwidth levels **all the way to a set of high performance end-host data transfer nodes (DTNs),**

2. Protocol agnostic traffic shaping services at the site edges and in the network core,

   Coupled to high throughput transfer applications providing stable, predictable data transfer rates

3. Machine learning + system modeling and Pervasive end-to-end monitoring

**To track, diagnose and optimize system operations on the fly**
Next Gen SDN Integrated Architecture

Summary

- The decadal challenges of globally distributed Exascale data and computation faced by major science programs must be addressed.
- New approaches: Deeply programmable software driven networked systems. We have just scratched the surface.
  - A new “Consistent Operations” paradigm: SDN-driven goal-oriented end-to-end operations, founded on:
    - Stable, resilient high throughput flows (e.g. FDT)
    - Controls at the network edges, and in the core
    - Real-time dynamic, adaptive operations among sites, the switched and optical network layers
    - Increasing negotiation, adaptation; built-in intelligence
    - Coordination among VO and Network Orchestrators
  - Bringing Exascale HPC and Web-scale cloud facilities, into the data intensive ecosystems of global science programs.
    - Petabyte transactions a driver of future network and server technology generations.