# SDN-NGenIA A Software Defined Next Generation integrated Architecture for HEP and Data Intensive Science

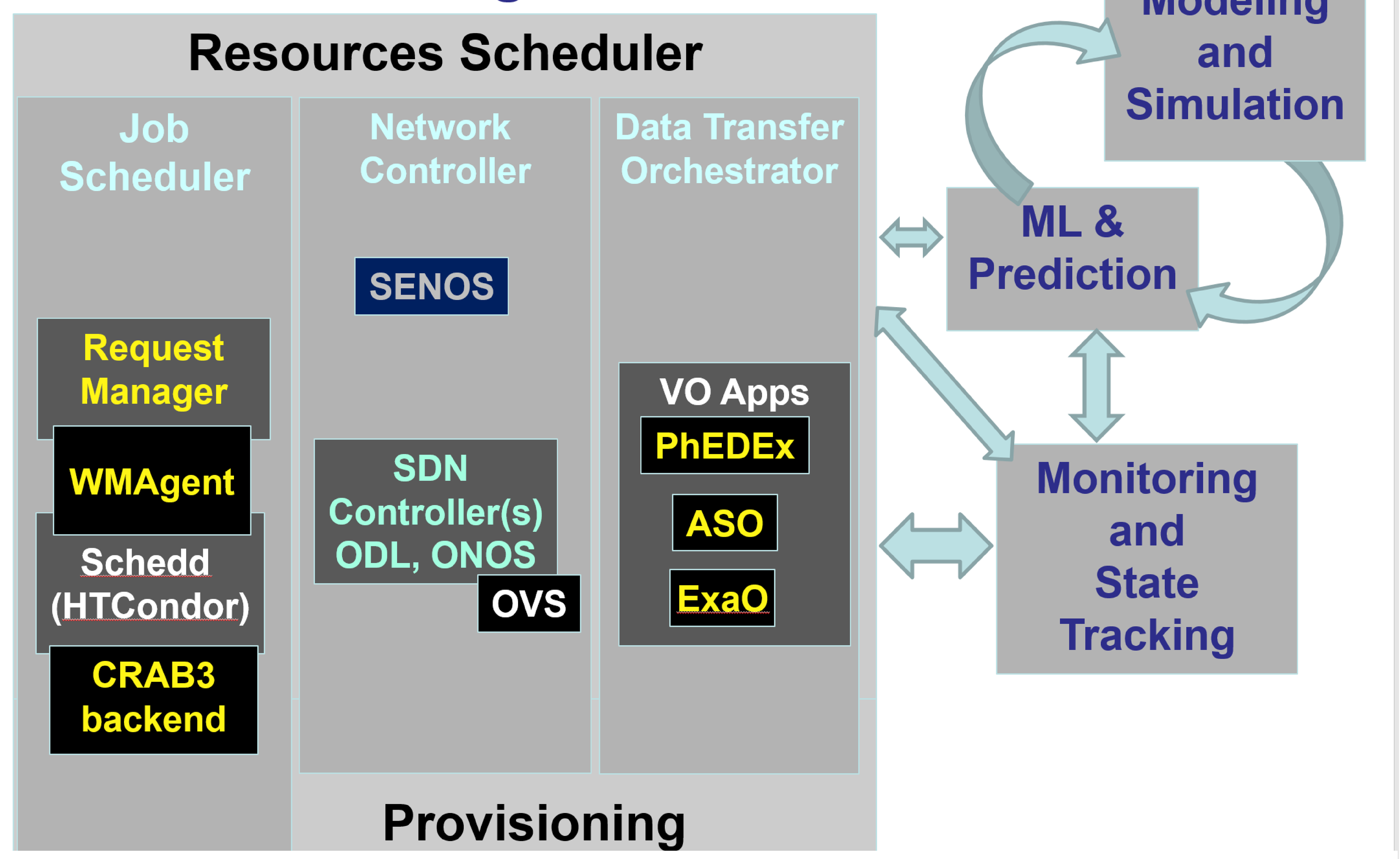
Dorian Kcira, dkcira@caltech.edu
California Institute of Technology

## Entering a New Era of Challenges as we Move to Exascale Data and Computing

- The largest science datasets today, from the LHC program are 300 petabytes
  - Exabyte datasets are on the horizon, by the end of Run2 in 2018
  - These datasets will grow by 100X, to the ~50-100 Exabyte range, during the HL LHC era from 2025
- The reliance on high performance networks will continue to grow as many Exabytes of data are distributed, processed and analyzed at hundreds of sites around the world
- As the needs of other fields continue to grow, HEP will face increasingly stiff competition for the use of large but limited network resources.



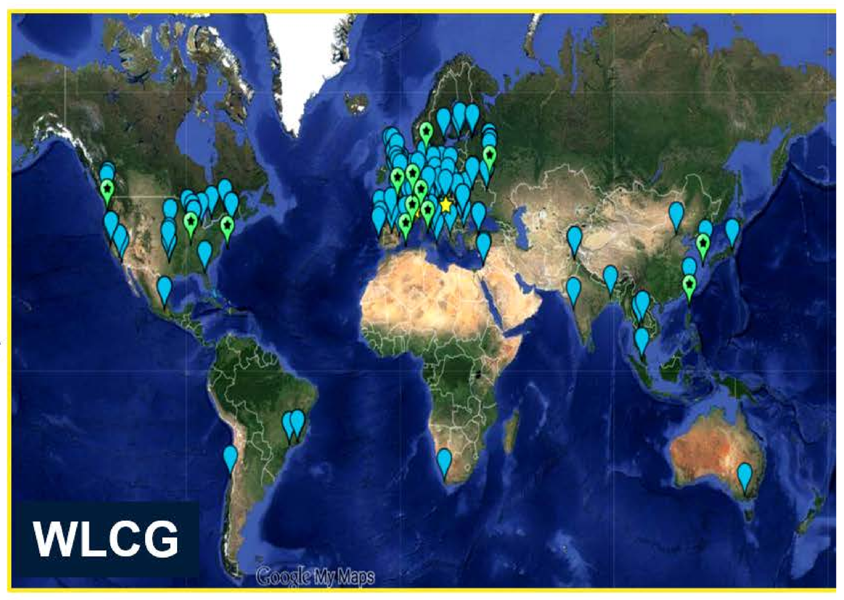

## Service Diagram: NGenIA




## Vision: Next Gen Integrated Systems for Exascale Science: Synergy ➡ a Major Opportunity

Exploit the Synergy among:

1. Global operations data and workflow management systems developed by HEP programs, being geared to work with *increasingly diverse and elastic resources to respond to peak demands*
   - Enabled by distributed operations and security infrastructures
   - Riding on high capacity (but mostly still-passive) networks




2. Deeply programmable, agile software-defined networks (SDN) Emerging as multi-domain network "operating systems"
   + New network paradigms focusing on content: from CDN to NDN
3. Machine Learning, modeling and simulation, and game theory methods Extract key variables; optimize; move to real-time self-optimizing workflows

The Watershed: A new ecosystem with ECFs as focal points in the global workflow; meeting otherwise daunting CPU needs

## SDN in SDN-NGenIA and SENSE

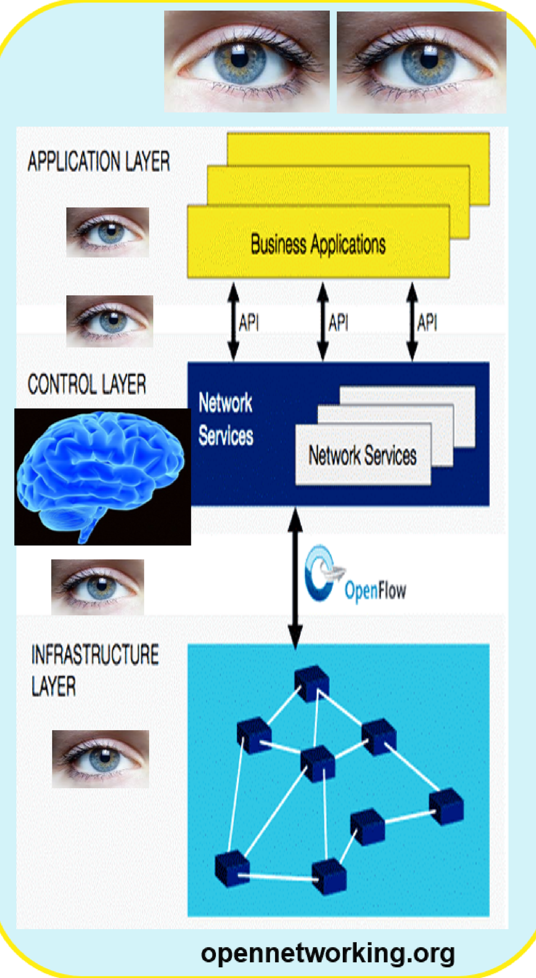
Building on Caltech/ESnet/FNAL Experience

*Vision:* Distributed environments where resources can be deployed flexibly to meet the demands

SDN is a natural path to this vision: separating the functions that control the flow of network traffic, from the switching infrastructure that forwards the traffic itself through open deeply programmable controllers.

With many benefits:

- Replacing stovepiped vendor HW/SW solutions by open platform-independent software services
- Virtualizing services and networks lowering cost and energy, with greater simplicity
- Enabling new methods and architectures
- A major direction of Research networks + Industry
- Still emerging and maturing

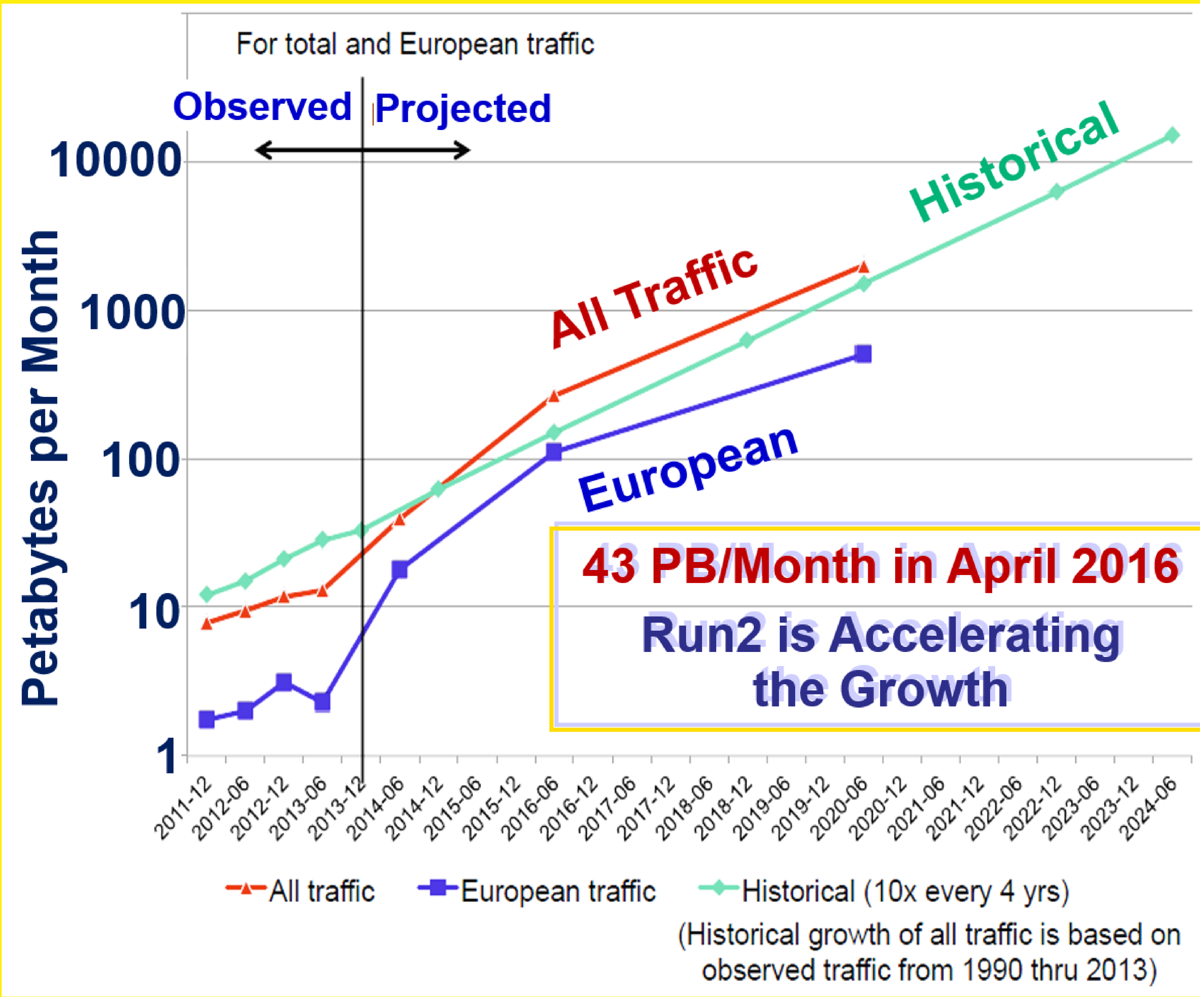



A system with built in intelligence
Requires excellent monitoring at all levels

## ESnet Science projection to 2024 Compared to historical traffic

E. Dart
W. Johnston

### Total traffic handled in Petabytes per Month




Projected Traffic Reaches 1 Exabyte Per Month by ~2020
10 EB/Mo. by ~2024

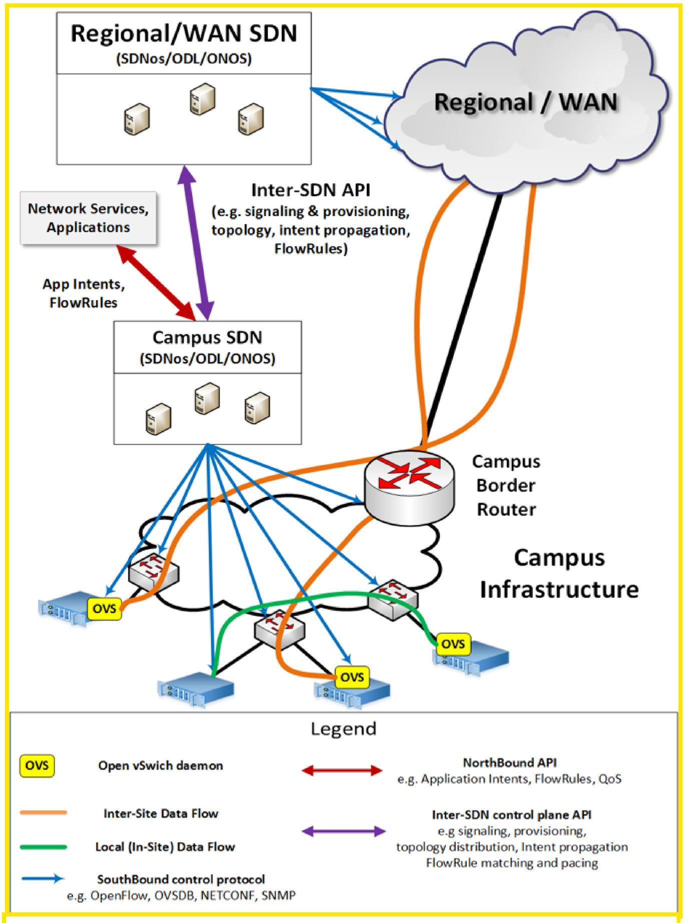
Rate of increase follows or exceeds Historical trend of 10X per 4 Years

HEP traffic will compete with BES, BER and ASCR

## OVS: Managing Site Interactions Locally, with Regional and Wide Area Networks

- Provides SDN-orchestrated configuration for data flows all the way to end-host, which can be orchestrated from the local/campus SDN controller or brought down from the Regional/WAN controller
- Provides QoS and traffic shaping right at the end-point of a data transfer
  - QoS via OVS is protocol agnostic: one can use TCP (GridFTP, FDT) or UDP
  - Helps to achieve better throughput by moderating and stabilizing data flows; e.g. in cases where the upstream switches have limited buffer memory
  - Under the hood, OVS uses the TC (Traffic control) part of iproute2 to configure and control the Linux kernel network scheduler
- Monitoring is done with standard sFlow and/or NetFlow protocols

www.openvswitch.org




OVS is included in the standard Linux releases

## The Future of Big Data Circa 2025: Astronomical or Genomical ? By the Numbers

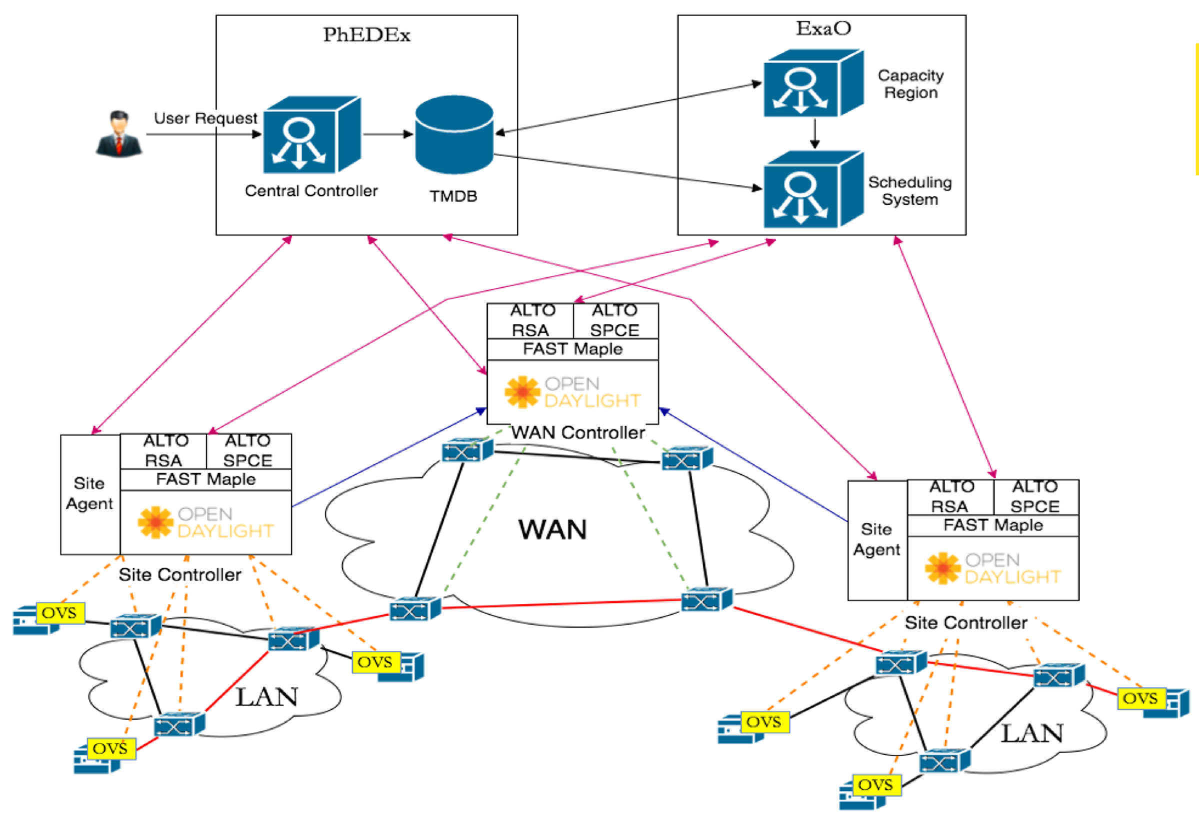
*PLoS Biol* 13(7): e1002195. doi:10.1371/journal.pbio.1002195

Domains of Big Data in 2025. In each, the projected annual and storage needs are presented, across the data lifecycle
Basis: 0.1 to 2B Humans with Genomes, replicated 30Xs;
+ Representative Samples of 2.5M Other Species' Genomes

| Data Phase | SKA | Twitter | YOU TUBE | GENOMICS | HL LHC |
|---|---|---|---|---|---|
| Acquisition | 25 ZB/Yr | 0.5–15 billion tweets/year | 500–900 million hours/year | 1 Zetta-bases/Yr | |
| Storage | 1.5 EB/Yr | 1–17 PB/year | 1–2 EB/year | 2-40 EB/Yr | 2-10 EB/Yr |
| Analysis | In situ data Reduction | Topic and sentiment mining | Limited requirements | | |
| | Real-time processing | Metadata analysis | | Variant Calling $2 \times 10^{12}$ CPU-h | |
| | Massive Volumes | | | All-pairs genome alignment $10^{16}$ CPU-h | 0.065 to 0.2 X $10^{12}$ CPU Hrs |
| Distribution | DAQ 600 TB/s | Small units of distribution | Major component of modern user's bandwidth (10 MB/s) | Many at 10 MBps Fewer at 10 TB/sec | DAQ to 10 TB/s Offline ~0.1 TB/s |

(1) Genomics Needs Realtime Filtering/Compression Before a Meaningful Comparison Can Be Made (2) New Knowledge is Transforming the Problem

## Consistent Network Operations in ODL




**Request and Data Management**

- Monitor users' requests
- Manage data locations
- Enforce data policies

**Network Abstraction and Service**

- Receives data requests, raw network state
- Compute on-demand, dynamic, inter-domain network abstraction
- Enforce application policies in networks

**Transfer Orchestrator**

- Compute and set scheduling point (notify site controllers)
- Learn and adapt

Y. Richard Yang
*Yale*

Site-resident: Site Agent, Site OpenDaylight Controllers

Network-resident: Abstraction (ALTO RSA) and Control (ALTO SPCE)

SciTools: Orchestrator/ Scheduling systems