



# Highlights from Track 7 - Facilities, Clouds & Containers

24th International Conference on Computing in High Energy & Nuclear Physics

4-8 November 2019, Adelaide, Australia

Oksana Shadura (Nebraska), Sang Un Ahn (KISTI), Christoph Wissing (DESY)

- > 48 oral presentations
  - 8 sessions – 2 of them in parallel on Thursday afternoon
- > ~22 posters
- > Typically 30-40 persons in the room
  - Track rather well attended
- > Good atmosphere
  - Often lively discussions – cut to stay (almost) on schedule

Pretty difficult to address all content in ~12min

Personal selection – it is biased!

Break									
Adelaide Convention Centre									
10:30 - 11:00									
11:00	Track 1 – Online and Real-time Computing: Hardware acceleration and hardware machine learning	Track 2 – Offline Computing: ML Tracking and parallelisatio	Track 3 – Middleware and Distributed Computing: Infrastructure & Identity	Track 4 – Data Organisation and Access: storage systems	Track 5 – Software Development Common tools 1: GUI, geometry, analysis, data models	Track 6 – Physics Analysis: Pheno fits / Analysis preservation	Track 7 – Facilities, Clouds and Containers: Non-LHC experiments	Track 8 – Collaboration, Education, Training and Outreach: Collaboratio and training	Track 9 – Exascale Science: Scheduling, computing environment
12:00	Hall G, Adelaide Convention Centre								
Lunch									
Adelaide Convention Centre									
12:30 - 14:00									
14:00	Track 1 – Online and Real-time Computing: Future upgrades	Track 2 – Offline Computing: Reconstructi and Performance	Track 3 – Middleware and Distributed Computing: Information Systems, BOINC & User Analysis	Track 4 – Data Organisation and Access: storage systems evolution and cha...	Track 5 – Software Development Common tools 2: Messaging, declarative frameworks, optimizers	Track 6 – Physics Analysis: Even Reconstruct and Selection methods	Track 7 – Facilities, Clouds and Containers: Opportunistic resources	Track 7 – Facilities, Clouds and Containers: Network technologies	Track 9 – Ex scale Science: Software environment, quantum algorithms, others
15:00									



Monitoring & Benchmarking



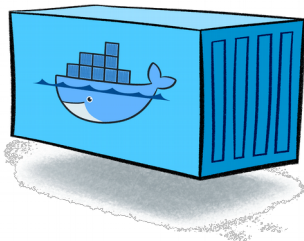
Cloud Computing



Trends & new Approaches



Infrastructure



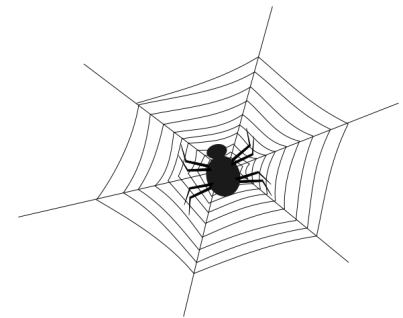
Containers



Non LHC Experiments



Opportunistic Resources



Network Technology

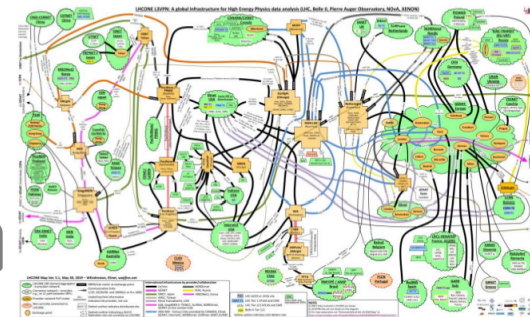
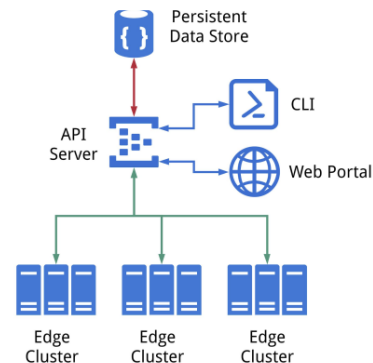
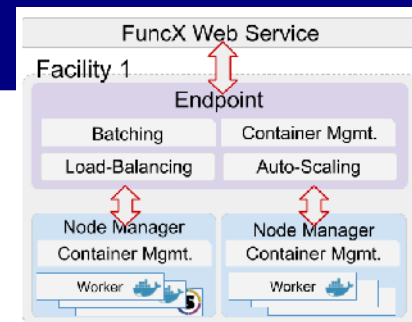
- > HEPspec06 benchmark does not correlate with real HEP workloads of today
  - Real reference HEP tasks can serve as new benchmark
  - Recent container technology enables practical packaging and distribution
- > Machine Learning techniques allow novel approaches for Monitoring
  - Anomaly detection via unsupervised networks
  - Inputs either metrics or logs
- > SAND network monitoring project
  - Gather metrics from various sources, e.g. perfSONAR, FTS, network workflows
  - Employ modern analytics tools
- > Common platform for monitoring: Elastic ecosystem
  - Used for MONiT at CERN, but also elsewhere



- > Commercial and research Clouds remain important
- > European Open Science Cloud (EOSC)
  - DODAS (Dynamic On Demand Analysis Service) using EGI FedCloud
  - CS3-APIs: Vendor independent layer for sync-and-share
  - Pre-commercial cloud resource procurement projects
- > Projects with commercial partners pose new challenges
  - Many of them are not technical
- > Commercial clouds: Data import, export and access
  - Providers appear to have very strong internal network fabric and also good WAN connectivity
  - Import is always for free
  - Hosting, accessing and exporting data comes with significant costs

# New Approaches

- > Function as a service – funcX
  - Provide access to functionality/calculation via webservice
  - Prototype with HEP analysis example shows nice scaling
- > Reduce efforts to run/provide site services
  - SIMPLE
    - > Easy deployment of middleware components from minimal config files
    - > Builds on Docker containers and Puppet
  - SLATE (Services Layer at the Edge)
    - > Docker, Kubernetes, and Helm to package and deploy services
    - > Central server to mediate requests being sent to participating edge Kubernetes clusters
- > Noted (Network Optimized Transfer of Experimental Data)
  - On demand configuration of redundant network links
  - Employs Rucio, FTS and involved SDNs



Fabiola

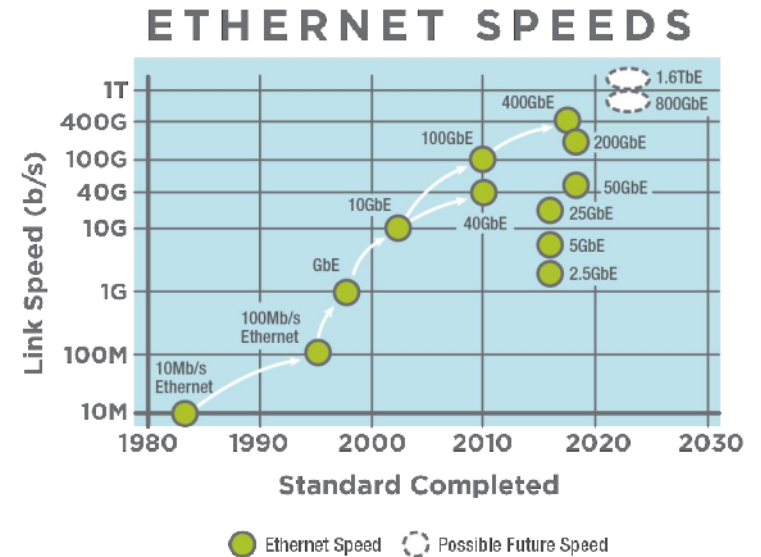


## > Jupyter

- Increasing interest on the user side
  - > Interactive, fast learning curve
  - > Easy development and sharing
- Jupyterhub enables access to diverse HTP and HPC resources

## > Overview by Hepix TechWatch working group

- Hyperscales (Google, Amazon) drive the market
- x86 market: AMD is back
- Magnetic disk: Market is shrinking
- Tape: Risks – essentially one company left for R&D
- Ethernet evolving very fast
  - > Pace of change exceeding IEEE standards process



## > Software infrastructure

- WPAD (Web Proxy Auto Discovery)
  - > Mechanism to easily auto setup Squid configuration for non-static resources
- Modernization of the ATLAS online web service environment

## > Site and hardware infrastructure

- Everlasting struggle with space and power
  - > CERN: After stopping extension at Wigner leverages containers from LHCb
  - > BNL: Setup of a new data center
- Challenges for locations with several sciences on campus
  - > Different demands from different communities



- HPC Cluster
- Users: Photon/Accelerator Scientists



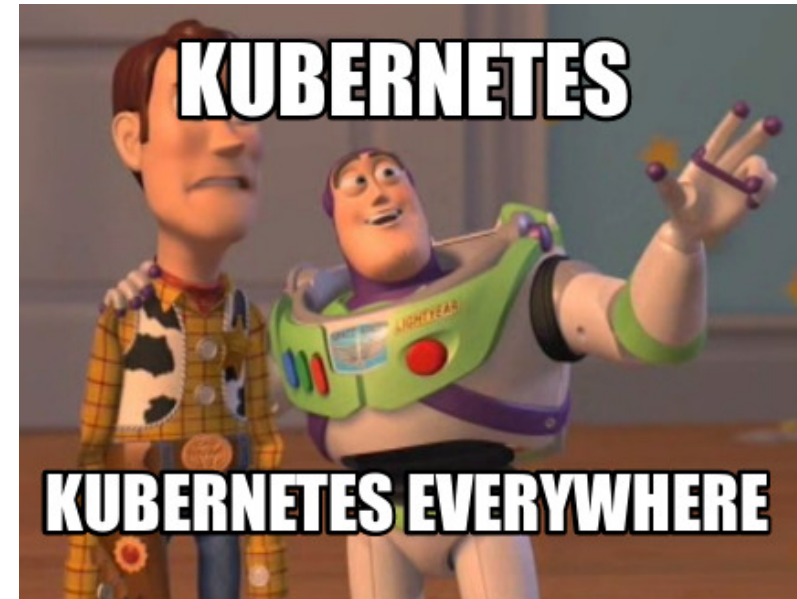
- HTC Cluster
- Users: HEP Scientists

- Difficult to understand and use one system well
- Do you want to enforce learning a second system?



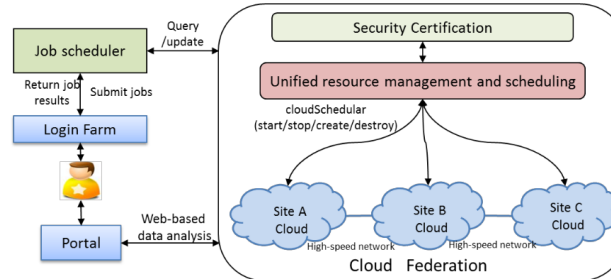


- > Kubernetes, kubernetes, kubernetes, ...
  - Intensive deployment of containers based on Kubernetes engine
  - Job scheduling without batch system, simply relying on Kubernetes  
-> effective and promising, aspect to simplifying site operations, no CEs...
  - Registry solutions to deploy container images is one of concerns
- > ScienceBox
  - Complete solution for scientific set of services from highly-scalable storage solutions (EOS) to user-friendly application, Jupyter notebook
  - All nicely packaged in containers
- > Container technology facilitates use of various resources
  - HPC, HTC, Grid resources, etc.
- > Moving CERN batch from Openstack VMs to Kubernetes
  - First benchmarks indicate 5% performance gain
- > All major experiments use containers in production



## > Leveraging resource access via Cloud interfaces

- LHAASO distributed computing
  - > Large High Altitude Air Shower Observatory
  - > cloudScheduler for multi Cloud access
- IceCube real-time processing in AWS



## > Non-LHC Communities building on LHC tools (or at least consider them)

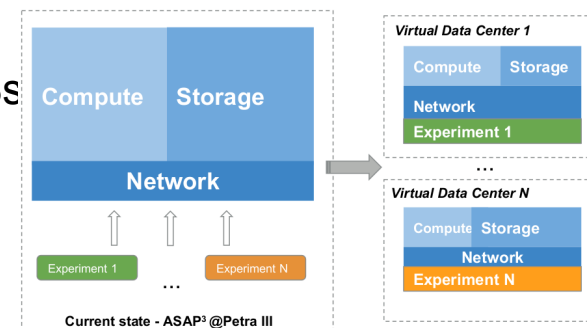
- RUCIO, FTS, HTCondor, CVMFS – very common for LHC groups
- VIRGO and Gravitational Waves computing in Europe
- AENEAS: Designing a Federated Regional Centre for SKA Computing
  - > Minimum storage: **750PB** over 10 years
  - > Several Exabytes after 15 years



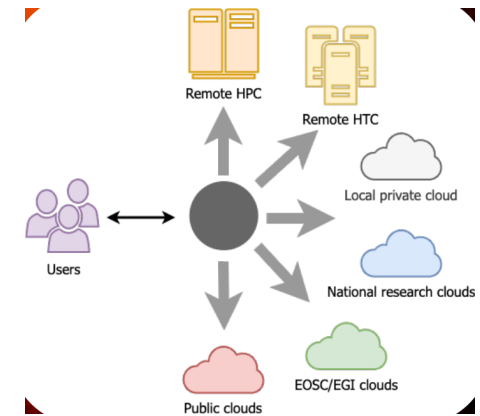
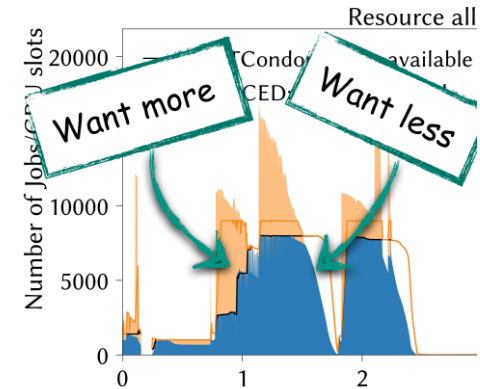
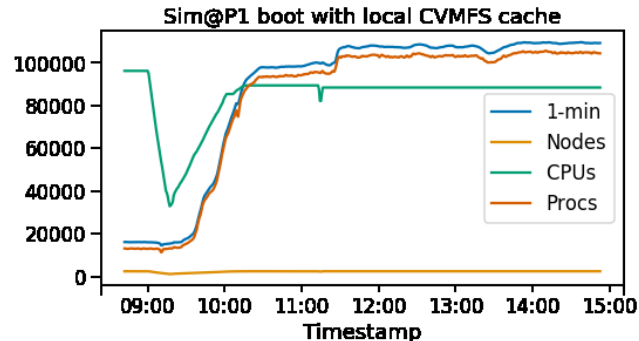
## > Smaller experiments (e.g. Project 8 and ADMX) move to tools like CI & K8s

## > Online computing for new generation photon science experiments

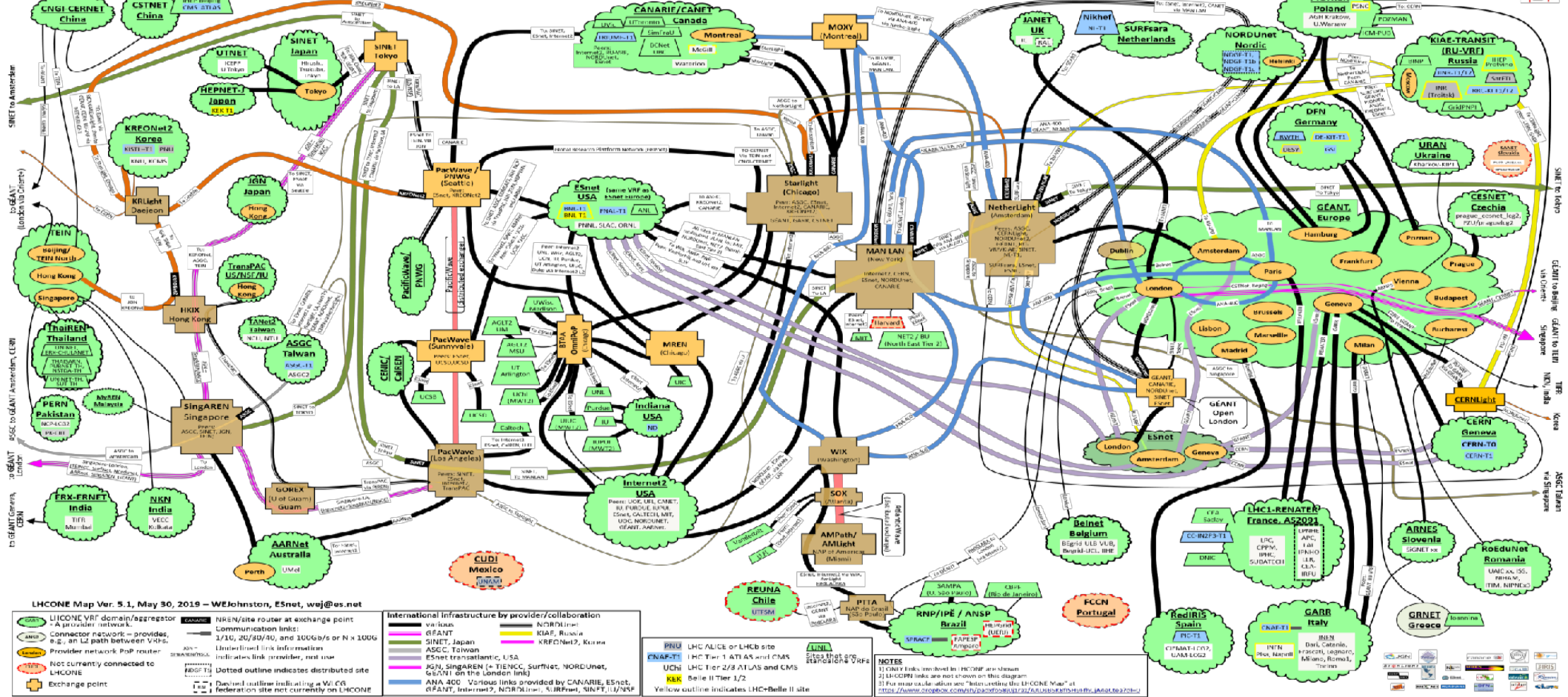
- Diverse requirements from various groups
- Modern Cloud stack allow to build virtual computing centers
  - > Isolate one group from another



- Adding academic or commercial Clouds and HPCs very common
- Different tools presented (there are more!)
  - COBaID/TARDIS from KIT
  - Google Cloud Platform Condor Pool Manager (GCPM) from Tokyo
  - GlueX (JLAB) utilizes resources at NERSC, PSU and OSG
- PROMINCE tool used for the fusion community (ITER)
  - Very flexible tool to access a spectrum of resources
  - “Users don’t need to worry about provisioning infrastructure on clouds
    - Or even know what a cloud is”
- Using HLT farms for offline processing
  - Common approach for LHC experiments
  - Recent improvements by ATLAS
    - Shorten the startup of offline environment
    - Persistent CVMFS cache



**LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON)**



**LHCONE Map Ver. 5.1, May 30, 2019 – WJJohnston, ESnet, waj@es.net**

**LHCONE VRF domain/aggregator**  
 - A provider network  
 - Connector network - provides, e.g., an IP path between VRFs  
 - Provider network PoB router  
 - Not currently connected to LHCONE  
 - Dotted outline indicates distributed site  
 - Dashed outline indicating a VRF  
 - Federation site not currently on LHCONE

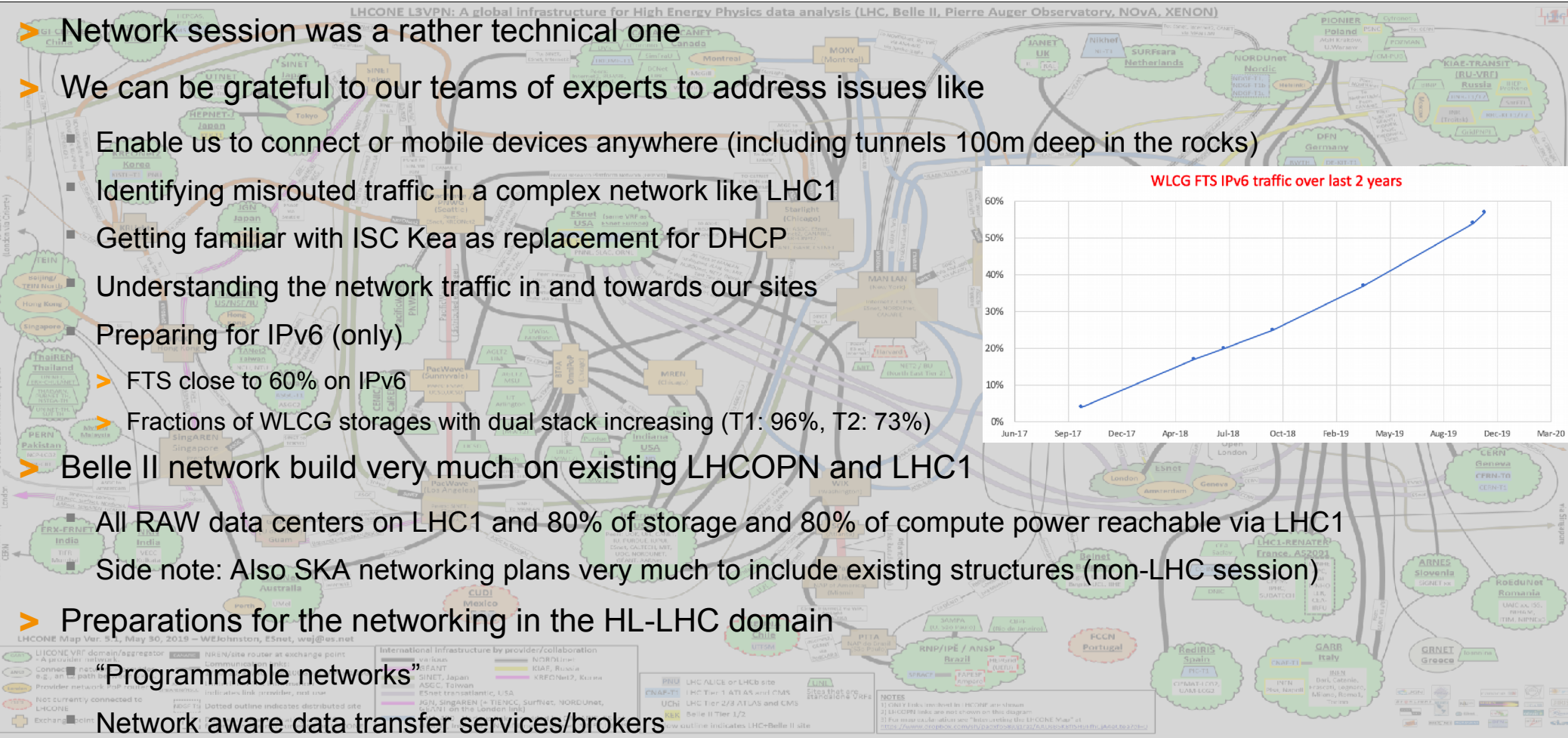
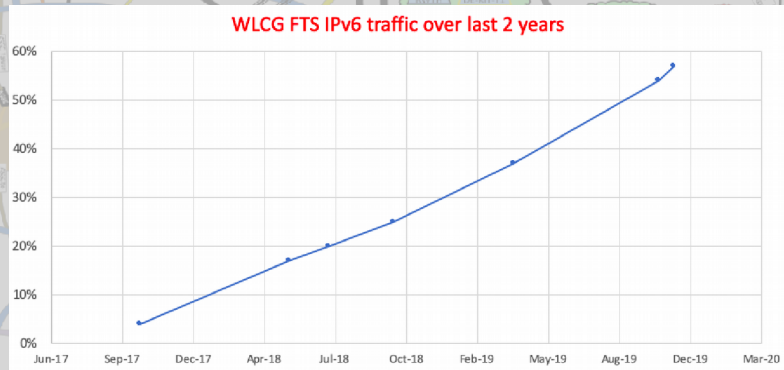
**Exchange point**

**International infrastructure by provider/collaboration**  
 - gFANT, ACTRIE  
 - SINET, Japan  
 - ASGC, Taiwan  
 - ESnet transatlantic, USA  
 - JGU, Singaren (+ TIENCC, SURFnet, NORDnet, G6/NI) on the London link  
 - ANA, ASC Various links provided by CANARIE, ESnet, gFANT, Internet2, NORDnet, SURFara, SINET, IU/NXP  
 - KRIAE, Russia  
 - KRFONet2, Korea  
 - UCI, LHC Tier 2/3 ATLAS and CMS  
 - Belle II Tier 1/2  
 - Sites that are transition VRFs

**NOTES**  
 1) CERN links involved in HCORP are shown  
 2) LHCORP links are not shown on this diagram  
 3) For more details about our "dot-coding" the LHCONE Map of CHEP 2019, please refer to the "dot-coding" section of the CHEP 2019 proceedings.

- > Network session was a rather technical one
- > We can be grateful to our teams of experts to address issues like
  - Enable us to connect or mobile devices anywhere (including tunnels 100m deep in the rocks)
  - Identifying misrouted traffic in a complex network like LHC1
  - Getting familiar with ISC Kea as replacement for DHCP
  - Understanding the network traffic in and towards our sites
  - Preparing for IPv6 (only)
    - > FTS close to 60% on IPv6
    - > Fractions of WLCG storages with dual stack increasing (T1: 96%, T2: 73%)
- > Belle II network build very much on existing LHCOPN and LHC1
  - All RAW data centers on LHC1 and 80% of storage and 80% of compute power reachable via LHC1
- Side note: Also SKA networking plans very much to include existing structures (non-LHC session)
- > Preparations for the networking in the HL-LHC domain
- “Programmable networks”
- Network aware data transfer services/brokers

LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOVA, XENON)



# And many more nice posters ... complementing the talks



## Setup and commissioning of a high-throughput analysis cluster

Rene Caspar<sup>†</sup>, Florian von Cube, Max Fischer, Manuel Giffels, Christoph Heidecker, Andreas Heiss, Eileen Kühn, Andreas Petzold, Günter Quast

**Introduction**

- Goals
  - Set up an analysis focused Tier-1 facility
  - Profit from existing Tier-1 infrastructure
  - Cluster optimized for usage with caching approaches
  - Throughput Optimized Analysis

**Setup of the Cluster**

- 11 hyperconvergent worknodes
- 1 TB NVMe and 96 TB HDD for local storage
- 100 Gbit/s network connection
- 1 GPU-node with 3 Nvidia V100

**Benchmarks**

- Testing analysis-like workflow
- Reading data from a ROOT file
- Typical speed limitation: ~50 MB/s
- Benchmark with up to 560 cores
- The benchmark is performed for:
  - Using NVMe SSDs
  - Using HDDs with CEPHFS as file system

**Usage of the TOPAS Cluster**

- Institute resources and development
- High-throughput extension of the cluster
- User jobs are forked to the TOPAS cluster for caching
- Distributed caches
- Hierarchical caches
  - See poster S10 by Max Fischer

**GPU use cases the BRIC experiment (Courtesy of H. Heidecker)**

- High-throughput analysis of BRIC experiment
- High-throughput analysis of BRIC experiment
- High-throughput analysis of BRIC experiment



## A Lightweight Door into the Grid

M. Maresch, J. M. Dorst, F. Wuerthwein, J. Sillig, J. Letts, E. P. Henninger, B. Liu, M. Seifried, D. Wetzel, B. P. B. Brauer, C. S. G. University of Chicago, University of Wisconsin-Madison, Argonne National Laboratory, Los Alamos National Laboratory, SLAC National Accelerator Laboratory, University of California, Berkeley

**Motivation**

As the amount of data collected in high energy physics experiments grows, providing an easy way to share their data and analysis is becoming a challenge. The Open Science Grid Compute Element (CE), the Grid portal to a Cloud-based CE initiative by the Open Science Grid, provides a lightweight solution to this problem.

**Hosted CEs**

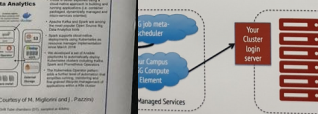
The Hosted CE is installed and managed by OSGrid staff. This offloads the responsibility of maintaining a CE to the site administrator for grid experts.



**Improvements**

Software and Operations teams have implemented the Hosted CE by implementing:

- Version control
- Configuration and hot fixes
- Software updates



**With the Hosted CE initiative, the hardware/software stack needed to operate a CE is maintained by OSGrid Operations staff in a homogeneous and automated way. This provides a reduction in the operational effort needed to maintain the CEs: one single organization does it in an uniform way, instead of each single resource provider doing it in their own way. Solutions that use containers to further reduce the operational footprint needed to maintain a set of CEs are being explored.**

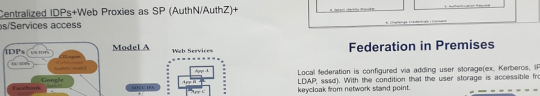
## Federated ID/SSO @BNL SDCC

Mizuki Karasawa, John Hoyer, Brookhaven National Laboratory

**Introduction**

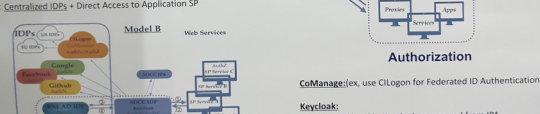
SDCC (Scientific Data and Computing Center) at BNL, now extended to a multi-tenant facility, a campus-wide scientific computing provider to many other institutions within the lab including NSLS-II, CFN, Biology etc. along with support for institutional projects that extend beyond BNL. The emerging needs for more robust and collaborative tool services require federated ID (externally or in-house) to be enabled in order to access protected resources. For external users, users will be able to access the resource at SDCC using their home institutional accounts, for internal federated access, users will be using their institutional accounts or lab-wide AD access for access.

**Centralized Vs. Distributed IDP models @SDCC**



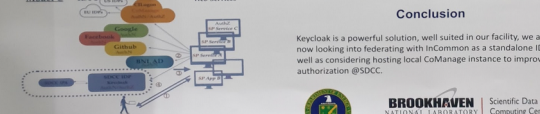
**Model A: Centralized IDPs + Web Proxies as SP (AuthN/AuthZ) + Apps/Services access**

Web Services: Kerberos, LDAP, SAML, OAuth, etc.



**Model C: Distributed IDPs + Direct Access to Application SP**

Web Services: Kerberos, LDAP, SAML, OAuth, etc.

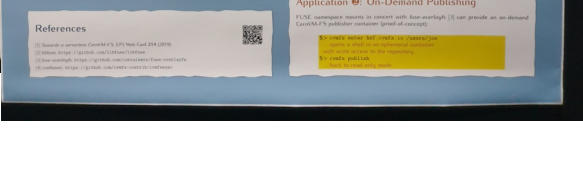
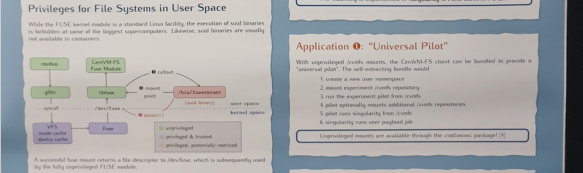
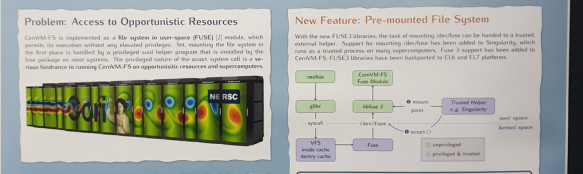
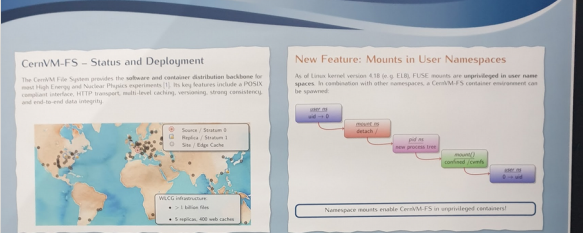


**Conclusion**

Keycloak is a powerful solution, well suited in our facility, we are now looking into federating with InCommon as a standalone IDP as well as considering hosting local CoManagement instance to improve the authorization @SDCC.

## A fully unprivileged CernVM-FS

J Blomer, D Dykstra, G Ganis, S Mosciatti, J Priessnitz, I CERN, \*Fermilab, jblomer@cern.ch



- > Our community migrates away from home grown tools towards industry standard products
  - Modern container solutions
  - Cloud architecture
  - HEP community is power user, sometimes contributor but almost never the main developer  
→ Poses some challenges in collaboration, typically less of technical nature
- > Many sites are no longer HEP only
  - Synergies with neighboring sciences: SKA and LIGO/VIRGO using HEP data management tools
  - Challenges: Dealing efficiently with a diverse user community

Closing the loop to the first plenary talk by H. Schellman:

**Do not pay too much attention to the summary speaker** (not a 1:1 quote)

- > Our community migrates away from home grown tools towards industry standard products
  - Modern container solutions
  - Cloud architecture
  - HEP community is power user, sometimes contributor but almost never the main developer  
→ Poses some challenges in collaboration, typically less of technical nature
- > Many sites are no longer HEP only
  - Synergies with neighboring sciences: SKA and LIGO/VIRGO using HEP data management tools
  - Challenges: Dealing efficiently with a diverse user community

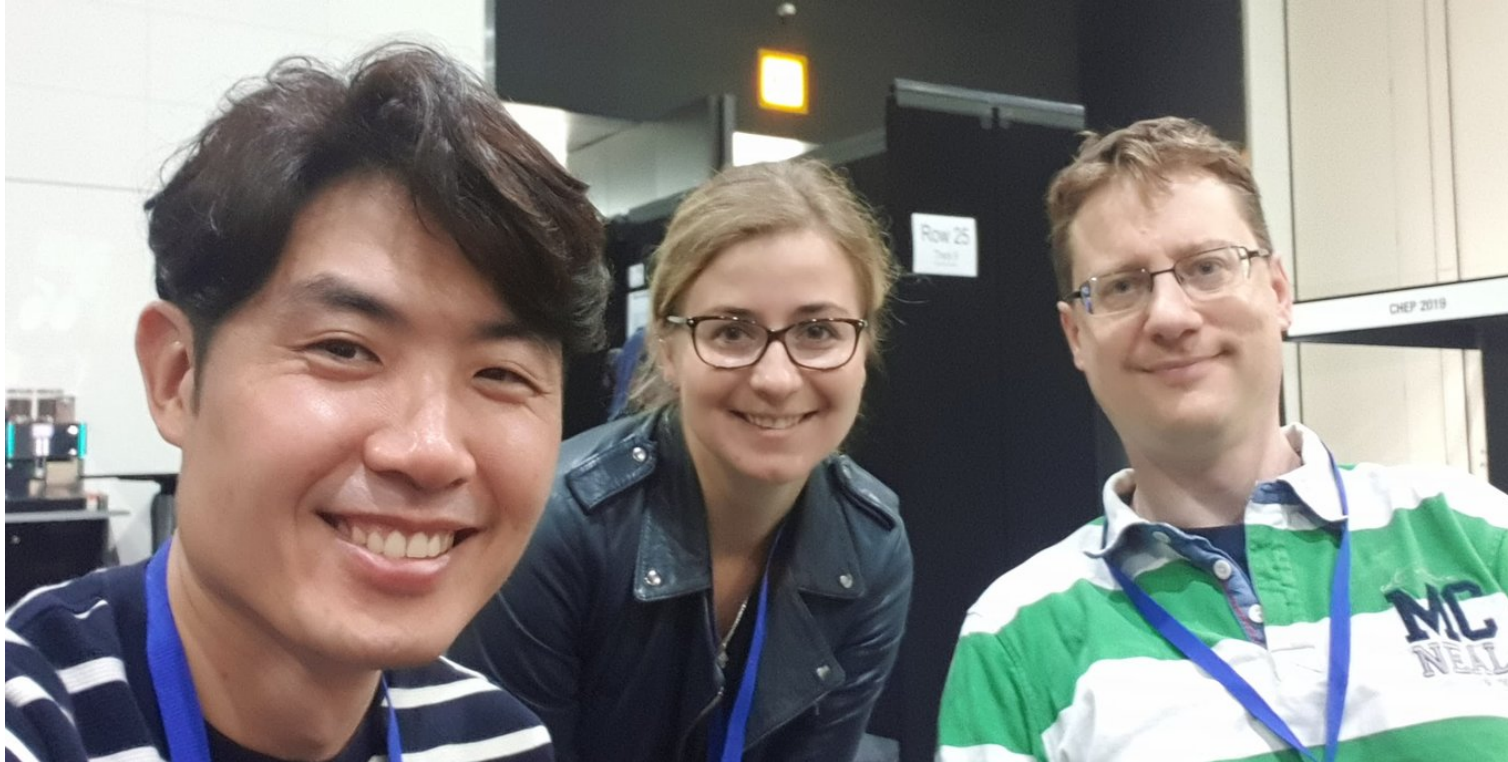
Closing the loop to the first plenary talk by H. Schellman:

**Do not pay too much attention to the summary speaker** (not a 1:1 quote)

...so instead of listing...



...just look at us saying:



THANKS a lot to all contributors to our track, the local organizers and the conference chair team