Network activities update

ALICE workshop
4th of May 2017
edoardo.martelli@cern.ch
Agenda

Updates on:
- Networks at CERN IT
- LHCOPN
- LHCONE
- IPv6 adoption
- Cloud Activities [if time allows]
The IT-CS group is responsible for communication services in use at CERN for data, control, voice and video

http://information-technology.web.cern.ch/about/organisation/communication-systems
Recent activities

Wigner remote data-centre:
- added 3rd 100G link Geneva-Budapest
- re-configured network to increase local capacity

Tender for new datacentre routers
- to replace Brocade MLXE contract
- on going, waiting for replies
- to be concluded by September
3rd 100G Geneva-Budapest link
Recent activities

Campus upgrades
- New WIFI network being deployed
- Planning for routers and switches upgrade

2nd network hub for business continuity in Prevezzin
- Construction works almost completed
- Installation of network devices will start in the summer

New datacentre in Prevezzin
- evaluating feasibility and costs of a shared DAQ facility located in Prevezzin (FR)
Shared DAQ (TBC)
Possible network topology

- **LHCO/PN**
- **Prevessin Network Hub**
- **T0-T1s network**
- **T1s-T2s network**
- **Internet**
- **Cloud resources**
- **Geneva DC Computing / Storage**
- **Wigner DC Extra capacity**
- **Up to 40 Tbps DAQ lines**
- **Not existing today**
- **Construction under evaluation**

- **CMS**
- **ALICE**
- **ATLAS**
- **LHCb**

- **Under construction**
CERN external traffic – last 12 months

Average bandwidth increased of ~80% compared to previous year
CERN external traffic – last 12 months

Peak of 250Gbps outbound
LHCOPN

Private network connecting Tier0 and Tier1s
- Reserved to LHC data transfers and analysis
- Single and bundled long distance 10G and 100G Ethernet links
- Star topology
- BGP routing: communities for traffic engineering, load balancing
- Security: only declared IP prefixes can exchange traffic.

https://twiki.cern.ch/twiki/bin/view/LHCOPN/WebHome
Latest developments

IPv6 adoption:
- 10 Tier1s and the Tier0 now have IPv6 connectivity
- dual-stack perfSONAR installed in all of them

Doubled link capacity during last year:
- NL-T1 (2x10G)
- FR-IN2P3 (2x10G)
- NDGF (2x10G)
- IT-INFIN-GARR (4x10G)
- UK-T1-RAL (2x10G, load balancing on existing backup link, now procuring 3rd 10G link)

CA-TRIUMF: is moving to SFU; TRIUMF dismantled by 2020
Complete IPv6 deployment:
- connect all Tier1s
- make all LHCOPN perfSONAR probes dual-stack
- use IPv6 for production data transfers

Upgrade Tier0-Tier1 links when necessary
- add additional 10G or move to 100G when cost effective
LHCONE L3VPN: Routed Virtual Private Network
- Worldwide private network connecting WLCG and other HEP sites at high bandwidth
- Bandwidth dedicated to HEP data transfers, no contention with other research projects
- Trusted traffic that can bypass slow perimeter firewalls

https://twiki.cern.ch/twiki/bin/view/LHCONE/WebHome
LHCONE L3VPN - status

- Over 20 national and international Research Networks

- Interconnections at Open Exchange Points including NetherLight, StarLight, MANLAN, WIX, CERNlight, Hong Kong and others

- Trans-Atlantic connectivity provided by ESnet, GEANT, Internet2, NORDUnet and SURFnet
- Trans-Pacific connectivity provided by ASGCnet, KREOnet, SINET
- Euro-Asia connectivity provided by GEANT, TEIN, TIFR

- ~74 end sites connected:
  - 14 Tier0/1s
  - 60 Tier2s
Traffic within LHCONE is steadily growing
- GÉANT has seen peaks of over 100Gbps
- Growth of over 65% from Q2 2015 to Q2 2016
- ESnet sees more LHCONE traffic than LHCOPN

Some NRENs and sites need to upgrade their connection
- GÉANT is already working with the NRENs for the upgrades

Expected to see further increases after the upgrades
LHCONE L3VPN - update

The LHCONE network is expanding
- Ukrain, Russia and Belgium connected to GÉANT (Russian ban finally lifted)
- Brazil connected to GÉANT, Chile is interested to join
- TEIN (Asia) connected to GÉANT, provides transit to Thailand
LHCONE L3VPN - current topology

Asia

North America

Europe

South America
LHCONE development in Asia

SINET Japan ready to peer with TEIN in Hong Kong, but existing interconnection turned out to be just 1Gbps. Will wait till TEIN’s routers are upgrade in June 2017.

Other peerings with TEIN are waiting for the TEIN’s routers upgrade.

IHEP and CCNU interested to connect to LHCONE. Discussion started with CERNET and CSTnet. CSTnet has promised to implement LHCONE.
Upcoming developments

ThaiREN TH AS24475
TEIN Asia AS24490
CERNlight CH AS20641

SINET JP AS2907
ASGCNet TW AS24167

ERnet IN AS2697
CERNET CSTNet China

ESnet US AS293
KREONET KR AS17579

GEANT EU AS20965

International
National

LHCONE peering in place
Missing LHCONE peering
Upcoming LHCONE peering
TransPAC

Cooperative partnership among Indiana University, APAN, TEIN, JGN/NICT-Japan, NII-Japan, CERNET, and others.

Started in March 2015, running through Feb 2020

$4.8M over 5 years. Includes funding for circuits (100G), exchange points, application support/engagement, research

Increased engagement with the wider community through groups such as APAN, TEIN/Asi@Connect, GNA, GLIF, LHCONE, Internet2, etc

Offers capacity over transpacific 100G link to LHCONE, to connect TEIN with Internet2 and ESnet
LHCONE connectivity via TransPAC
ALICE sites connected to LHCONE

ALICE sites:
~ 75 sites around the world
- at least 35 sites connected to LHCONE
<table>
<thead>
<tr>
<th>Site</th>
<th>Group</th>
<th>Tier</th>
<th>LHCONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vienna</td>
<td>Austria</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>CBPF</td>
<td>Brasil</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>SaoPaulo</td>
<td>Brasil</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>CERN (Meyrin)</td>
<td>CERN</td>
<td>T0</td>
<td>y</td>
</tr>
<tr>
<td>CERN (Wigner)</td>
<td>CERN</td>
<td>T0</td>
<td>y</td>
</tr>
<tr>
<td>Prague</td>
<td>Czech Republic</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>FZK</td>
<td>Germany</td>
<td>T1</td>
<td>y</td>
</tr>
<tr>
<td>GSI</td>
<td>Germany</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Athens</td>
<td>Greece</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>KFKI</td>
<td>Hungary</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>CCIN2P3</td>
<td>IN2P3</td>
<td>T1</td>
<td>y</td>
</tr>
<tr>
<td>Clermont</td>
<td>IN2P3</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Grenoble</td>
<td>IN2P3</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>GRIF_IPNO</td>
<td>IN2P3</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>GRIF_IRFU</td>
<td>IN2P3</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>IPNL</td>
<td>IN2P3</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>Strasbourg_IRES</td>
<td>IN2P3</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Subatech_CCIP</td>
<td>IN2P3</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Kolkata</td>
<td>India</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Bandung</td>
<td>Indonesia</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>Cibinong</td>
<td>Indonesia</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>Bari</td>
<td>INFN</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Cagliari</td>
<td>INFN</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>Catania</td>
<td>INFN</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>CNAF</td>
<td>INFN</td>
<td>T1</td>
<td>y</td>
</tr>
<tr>
<td>Site</td>
<td>Group</td>
<td>Tier</td>
<td>LHCONE</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Legnaro</td>
<td>INFN</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Torino</td>
<td>INFN</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Trieste</td>
<td>INFN</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>TriGrid_Catania</td>
<td>INFN</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Hiroshima</td>
<td>Japan</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>Tsukuba</td>
<td>Japan</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>UNAM</td>
<td>Mexico</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>CSC</td>
<td>Nordic Countries</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>DCSC_KU</td>
<td>Nordic Countries</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>HIP</td>
<td>Nordic Countries</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>LUNARC</td>
<td>Nordic Countries</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>SNIC</td>
<td>Nordic Countries</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>UIB</td>
<td>Nordic Countries</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>COMSATS</td>
<td>Pakistan</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>PAKGRID</td>
<td>Pakistan</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>Cyfronet</td>
<td>Poland</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Poznan</td>
<td>Poland</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>WUT</td>
<td>Poland</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>IHEP</td>
<td>RDIG</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>ITEP</td>
<td>RDIG</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>JINR</td>
<td>RDIG</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>MEPHI</td>
<td>RDIG</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>PNPI</td>
<td>RDIG</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>RRC-KI</td>
<td>RDIG</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>SPbSU</td>
<td>RDIG</td>
<td>T2</td>
<td></td>
</tr>
</tbody>
</table>
# ALICE sites

<table>
<thead>
<tr>
<th>Site</th>
<th>Group</th>
<th>Tier</th>
<th>LHCONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troitsk</td>
<td>RDIG</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>KISTI</td>
<td>Republic of Korea</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>ISS</td>
<td>Romania</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>NIHAM</td>
<td>Romania</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>NIPNE</td>
<td>Romania</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>UPB</td>
<td>Romania</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>Bratislava</td>
<td>Slovakia</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>Kosice</td>
<td>Slovakia</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>ZA_CHPC</td>
<td>South Africa</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>Madrid</td>
<td>Spain</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>Trujillo</td>
<td>Spain</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>NECTEC</td>
<td>Thailand</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>SUT</td>
<td>Thailand</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>NIKHEF</td>
<td>The Netherlands</td>
<td>T2</td>
<td>y</td>
</tr>
<tr>
<td>SARA</td>
<td>The Netherlands</td>
<td>T1</td>
<td>y</td>
</tr>
<tr>
<td>Birmingham</td>
<td>UK</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>Oxford</td>
<td>UK</td>
<td>T2</td>
<td>n</td>
</tr>
<tr>
<td>RAL</td>
<td>UK</td>
<td>T1</td>
<td>n</td>
</tr>
<tr>
<td>BITP</td>
<td>Ukraine</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>ICYB</td>
<td>Ukraine</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>ISMA</td>
<td>Ukraine</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>KNU</td>
<td>Ukraine</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>LBL</td>
<td>US</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>LLNL</td>
<td>US</td>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>ORNL</td>
<td>US</td>
<td>T2</td>
<td></td>
</tr>
</tbody>
</table>
WLCG perfSONAR

278 perfSONAR instances registered in GOCDB/OIM
249 Active perfSONAR instances
perfSONAR - update

Mesh have been reconfigured/optimized. There are now experiment specific meshes: ATLAS, CMS and LHCb which allow daily bandwidth tests. ALICE uses MonaLISA

All the LHCOPN and LHCONE data is available in ElasticSearch via OSG (ongoing)

Some degradation in the data retrieving of the LHCOPN/ONE MaDDash. Upcoming upgrade to v4 should be the opportunity to make the nodes more resilient
perfSONAR - update

ATLAS is working on getting network metrics into an analytics platform:
- packet loss from perfSONAR
- network stats and flows from ESnet routers
- stats from CERN routers will be added

Looking for more router sources

Completed MCA (Mesh Configuration Admin) tool. Could be used by WLCG

WLCG is working on ETF (Experiment Test Framework) to monitor perfSONAR services
Open to other HEP collaborations

The L3VPN is now used also by:

- Belle II experiment
- NOvA neutrino experiment
- Pierre Auger observatory
- XENON dark matter project
GNA - Global Network Architecture

Sharing of R&E network resources for common benefit

Extend benefits gained with ANA-300G experience
  (sharing of three 100G transatlantic links)

Pillars:
- resource sharing
- aligning operational standards
- maximizing funding and investments
- knowledge sharing and outreach
- increasing global collaboration
Possible change of computing model

“Consolidate storage in few locations to decrease costs”

Picture: Simone Campana, Ian Bird - WLCG
Networking for Run4

Proposed project to improve network utilization efficiency
IPv6 deployment
IPv6 adoption

http://6lab.cisco.com/stats/
HEPiX IPv6 working group

IPv6 deployment have been followed up by the HEPiX IPv6 working group

It surveys IPv6 compliance of HEP applications and IPv6 connectivity of HEP sites

Lists are maintained at the web site: http://hepix-ipv6.web.cern.ch/
IPv6 only worker nodes

IPv6 requirements for LHC Experiments have been discussed at the GDB.

A plan to support IPv6 only worker nodes has been proposed. The aims are:
- provide a migration path for sites
- allow long term planning for network setups
- give the possibility to make use of IPv6 only resources, should they become available

It is essential to provide dual-stack access to storage.
Dual-stack storage

Most storage solutions and protocols now work in dual-stack mode
- dCache, DPM, StoRM, FTS
- XrootD 4, GridFTP, WebDAV

Several sites have been running dual-stack for some time

Production data transfers over IPv6 have started happening
IPv6 at Tier0-Tier1s

- Good IPv6 adoption
- 10 Tier1s and the Tier0 peering over IPv6
- dual-stack perfSONAR installed in all of them

Still missing:
- KR-KISTI (work in progress)
- RRC-KI-T1 KIAE (work in progress)
- RRC-KI-T1 JINR (will follow KIAE)

No IPv6 storage in most of them though, including CERN
IPv6 at Tier2s

IPv6 connectivity provided by most of the LHCONE VRFs
12 Tier2s have IPv6 connectivity with LHCONE

HEPiX is writing an How-To deploy IPv6 at a Tier2:
https://hepix-ipv6.web.cern.ch/content/how-deploy-ipv6-wlkg-tier-2-site

An IPv6 tutorial is scheduled for the WLCG workshop in Manchester in June
IPv6 and ALICE

ALICE central services have been dual stack for more than a year

Storage is fully federated based on XrootD. XrootD 4.1 is needed to support IPv6

To support IPv6-only resources, all data must be available on some IPv6-enabled storage

ALICE can support IPv6-only CPU resources as soon as enough sites have upgraded their storage to dual stack. At the moment only 5% have done it. Aiming to all sites dual-stack by the end of Run II
IPv6 Monitoring

Network performance on perfSONAR

FTS transfers in MONIT
https://monit.cern.ch/
(add “data.ipv6:true” to any FTS monitor page for IPv6 only traffic)

ETF
https://etf-ipv6-dev.cern.ch/etf/check_mk/
Clouds activities
HELIX NEBULA the Science Cloud

Procurers: CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, SURFSara, STFC
- Procurers have committed funds (>1.6M€), manpower, use-cases with applications & data, in-house IT resources

Objective: procure innovative IaaS level cloud services
- Fully and seamlessly integrating commercial cloud (Iaas) resources with in-house resources and European e-Infrastructures
- To form a hybrid cloud platform for science

Services will be made available to end-users from many research communities: High-energy physics, astronomy, life sciences, neutron/photon sciences, long tail of science

Co-funded via H2020 (Jan’16-Jun’18) as a Pre-Commercial Procurement (PCP) project: Grant Agreement 687614, total procurement volume: >5M€
HNSciCloud challenges

Innovative IaaS level cloud services integrated with procurers in-house resources and public e-infrastructure to support a range of scientific workloads

**Compute and Storage**
support a range of virtual machine and container configurations including HPC working with datasets in the petabyte range

**Network Connectivity and Federated Identity Management**
provide high-end network capacity via GEANT for the whole platform with common identity and access management

**Service Payment Models**
explore a range of purchasing options to determine those most appropriate for the scientific application workloads to be deployed
High Level Architecture of the Hybrid Cloud Platform including the R&D challenges
Project phases

We are here

Preparation
- Analysis of requirements, current market offers and relevant standards
- Build stakeholder group
- Develop tender material

Jan’16

Implementation and sharing
- 4 Designs
- 3 Prototypes
- 2 Pilots

Tender Jul’16 Call-off Feb’17 Call-off Oct’17 Dec’18

Each step is competitive - only contractors that successfully complete the previous step can bid in the next
Selected bidders

Admitted to the Design Phase
- T-Systems, Huawei, Cyfronet, Divia
- IBM
- RHEA Group, T-Systems, exoscale, SixSq
- Indra, HPE, Advania, SixSq
Other major players not interested or dropped out just before tender submission

Admitted to the Prototype Phase
- T-Systems, Huawei, Cyfronet, Divia
- IBM
- RHEA Group, T-Systems, exoscale, SixSq

Current Status: Evaluation of Bids in the Prototype Phase
Prototype phase

Collect a number of tests from the Procurers
- To test/validate the design proposed for the several project challenges
- To be able to assess results of the tests as one of the main criteria to be eligible to the next phase (Pilot)

Tests consist in applications and some selected workloads

Next phase (Pilot) will be based on the Prototype to be assessed on 3 main aspects:
- Performance
- Scalability
- Security
CMS GoogleCloud engine demo at SC 2017

Challenge: Doubling CMS computing using Google Cloud Engine
Expand the Fermilab facility to an additional 160,000 cores
Use HEPCloud technology to do this as transparently as possible to the application

Live demo during Supercomputing 2016
730172 simulation jobs submitted; only 47 did not complete
205 M physics events generated, yielding 81.8 TB of data
Cost: ~1.6 cents per core-hour (on-premises: 0.9 cents per core-hour assuming 100% utilization)
Conclusion
Summary

CERN: planning upgrades for RUN3 and for better resiliency

LHCOPN: stable, almost fully dual-stack, upgraded when needed

LHCONE: expanding, works on-going in Asia

IPv6: good progress in WAN networks, less in campus and datacentres. More dual-stack storage needed in WLCG

Clouds: testing integration with orchestration tools to fit peaks of computing demand
Questions?

edoardo.martelli@cern.ch