

HEPIX 2017 Budapest Summary Report

Tony Wong (BNL)

Pete Gronbech (additions, thanks to Martin Bly)

About Wigner

- Wigner Research Centre for Physics, RCP belongs to Hungarian Academy of Sciences. RCP roots go back to 1950s, formally established 2012 with merger of two institutions.
- DC milestones – T2 centre on Campus in 2004, CERN T0 tender win in 2012, construction on 2012-13, operation started in Jan 2013.
- Also value-added Cloud services to the academic sphere in Hungary: Wigner Cloud and Academic Cloud.

Statistics

- First HEPIX meeting @ Budapest
- 125(!) registered participants
 - 102 from Europe, 13 from Asia, 9 from North America and 1 from Israel
 - Representing 35+ organizations
- 66 contributed presentations
- Special network security workshop by Liviu Valsan (CERN)
 - A call to coordinate response to security incidents among HEPIX sites

More Statistics

- Presentations by track
 - Site Report (18)
 - Security & Network (9)
 - Storage & Filesystems (8)
 - Grid & Clouds (7)
 - Computing & Batch (8)
 - IT Facilities & Business Continuity (4)
 - Basic IT Services (8)
 - End-User Services (2)
 - Miscellaneous (2)

Site Reports

- CERN's
 - Migration away from AFS underway
 - Some impact at various sites
 - Replacement options (CERN Box, BNL Box, IHEP Box, EOS, CVMFS, etc)
 - CERN tape archive is 109PB
 - Migrating from LSF to Condor for all LHC users
 - ~30k VMs
 - Use Puppet 4
- Bologna, DESY, LAL, BNL, RAL, etc reported on the construction of new facilities and activities
 - Absorb smaller data centers and consolidate for more efficient operations
 - Branch out into newer domains (ie, weather forecasting & atmospheric research, photon sciences, biology, etc)

Site Reports

- Increasing adoption of CentOS/SL7 (LAL, etc), Ceph (LAL, BNL, RAL, KISTI, etc)
- Migration to Grafana (RAL, BNL, CSC, etc)
- PUE reduction at RAL to 1.35 after cooling upgrade
- Continuing presence by non-traditional fields at HEPIX (MDC, Alba, etc) is a good opportunity for synergy and interactions in new areas (HPC, Medical, etc)
- Multiple (unplanned) power outages at NDGF—UPS back-up operating as expected but still affected resource availability
- Increasing adoption of IPv6 dual-stack at many sites
- KEK estimates BELLE-II will start in early 2018—heavy load for computing facility (growth in usage of $\sim 1.6x$ HS06 over last year)

Site Reports

- Tokyo ATLAS Tier2 and IHEP migration to HTCondor—steady trend among many HEP sites
- KISTI data center supports LHC experiments, LIGO, BELLE and others – evaluating Kubernetes for container management
- Growing support for HPC activities within community (IHEP, LAL, BNL, etc) – opportunities for synergy and collaboration

End-User IT Services

- CERN presentation on Linux support described the CentOS7 effort – building infrastructure for testing, QA and release
- SW and Computing Journal presentation by Michel – opportunity for HEPIX community to publish selected contributions and share with other fields → avoid duplication of effort to address similar needs

Security & Networking

- Gradual increase in availability of IPv6 dual-stack among sites (RAL, IHEP, PIC, etc)
- Liviu presented a summary of current security compromises at computing sites
 - responses to compromises show lack of coordination, policy coherence, trust, forensics protocols, etc
 - Email is still favorite tool for hackers (seems to be universally true)
- Shawn reported on the WLCG Network Throughput WG. Some of the activities/goals are:
 - perfSONAR 4.1 coming—will require new deployment models for OSG/WLCG
 - Use monitoring data for notification/alerts, predictive analysis, usage metrics and commercial cloud performance

Security & Networking

- ESnet tasked with providing connectivity for DOE Science—ESnet6 covers 2020-2025 (possibly 2030)
 - New developments (clouds, scale of data movement, computing models, etc) make estimating network needs challenging
 - Closer coordination between ESnet and LHC community is desirable to improve efficiency of network upgrades and deployment
- IHEP's discussion of SDN-based network security—optimize algorithms to detect anomalies
- WLCG deployment plan presented a timeline for deployment of IPv6 (cpu, storage services, etc), along with status across LHC experiments, Tier 0/1/2 sites and remaining issues—concern about slow progress at Tier 2 sites

Security & Networking

- KEK report on persistent cyber attacks (phishing, viruses, etc) and the social engineering challenge → continuous need for user education
- Presentation on CERN's security operations center → detection, containment and remediation of security threats

Basic IT Services

- Jerome's presentation on migration from Puppet 3 to 4 → mostly smooth with a few hic-ups (user crontab and compilation times) → useful experience for future migration to Puppet 5
- Owen Synge described Salt as an alternative to Puppet, Chef and Ansible. Pros and cons of all solutions discussed
- Centralizing Elasticsearch @ CERN – discussion of project goals and current status
 - Challenges bigger than expected (ie, variety of use cases and requirements)
 - Deployment of ACL's and automation of workflows on-going
- Evolution of monitoring at CNAF
 - Cloud-oriented to match distributed nature of resources
 - Common infrastructure but custom environment for each site—future integration with ELK

Basic IT services

- Consolidation and unify CERN IT monitoring
 - Treat data sources separately—too diverse
 - Data Center monitoring migrating to Collectd
- LBL's data collection and comprehensive monitoring—160 GB and 1.2 billion docs/day!
- Presentation by Balabit (Hungarian company) on Syslog-ng following by Fabien's real case usage at IN2P3
 - Flexible, fast, small footprint and support are considered a plus

Storage & Filesystems

- CERN presented a strategic IT outlook for the future – building a flexible infrastructure solution at the exabyte scale
 - EOS+CTA to replace CASTOR (same tape format) by end of 2018
 - CTA available outside CERN (with some effort)
- CERNBox provides user access to EOS
 - Currently 9500+ users and 1.1 PB of used (3.3 PB deployed) storage
 - Generic HOME directory to replace AFS
- BNLBox is Ceph-based and targeted at BNL users
 - User access to HPSS tape storage
 - Expect 7.5 PB available by end of 2017
- Presentation by Seagate on current storage technologies
 - New techniques to increase capacity per drive
 - Constraints in IOPs/TB with increasing capacities

Storage & Filesystems

- NRC presentation on federated data storage technologies using geographically diverse centers within Russia
 - Performance evaluation of EOS and dCache-based solution using synthetic tools (ie, Bonnie++) and perfSONAR
 - Promising results but tests on-going
- RAL's CVMFS (Stratum-0 and 1) service provided to geographically-diverse sites (some outside of Europe)
 - “secure” CVMFS service to user communities with access-restricted (ie, OSG security credentials) requirements
- RAL also presented on its experiences with Ceph-based storage (production usage and incidents)
 - ECHO to provide disk-only storage and replace CASTOR for LHC VO's
- LAL presentation on Ceph experiences across its various programs (Agata, Cloud@VD, etc)
 - Distributed management team to provide common solution to all supported programs

Computing & Batch Services

- CosmoHub created to share data among cosmology projects
 - Migrating from PostgreSQL to Apache Hadoop+Hive to improve scalability and data analysis response time
- HammerCloud
 - WLCG testbed → essential tests and automation tool
 - 180k jobs daily
 - Data Center commissioning → job submission to resources behind HTCondor-CE
- Knights Landing (KNL) clusters
 - BNL installed a 144 test cluster
 - Many surprises and delays – hope to enter production in May 2017
 - ATLAS benchmarked BNL KNL cluster – good performance with optimization
 - JLAB cluster a 192 production cluster
 - Similar problems
 - Currently in production with LQCD community

Computing & Batch Services

- Report from Benchmarking WG
 - Discussion about detailed study of fast benchmark and replacing HS06 for guidance on procurement and site pledge. Various candidates (ATLAS KV, DB12, ROOT Stress test, etc)
 - Start requirements to validate successor to HS06
 - Must be representative of WLCG job mix
 - Must disentangle several effects (HT on/off, bare metal vs. VM, OS versions, etc)
- Separate presentation on DB12 to study discrepancy in results between Ivybridge and Haswell
- IHEP migration to HTCondor also includes unifying resources and allow for cross-usage → goal is to increase overall utilization
- Singularity presentation by Brian (Nebraska)
 - Job isolation and traceability within a container
 - Alternative to glxec
 - Being tested at various sites (as alternative to Docker)

Grids, Clouds & Virtualization

- CRIC (Computer Resources Information Catalog) is the follow-up to AGIS (ATLAS Grid Information System)
 - Integrates resource status and configuration with experimental framework
 - Unlike AGIS, CRIC is geared for WLCG community and not experiment-specific—however requires effort by local sites to participate
- ElastiCluster → set of tools to manage compute resources hosted on IaaS cloud infrastructure
 - Used for educational, testing and scaling purposes
- CERN update on Openstack-based cloud services
 - Adding 86k cores
 - Adding new components (Ironic for baremetal, Manila for file sharing, Magnum for software distribution, etc)

Grids, Clouds & Virtualization

- Ian's presentation on simplifying access to public clouds @ RAL
 - Use Kubernetes instead of vendor API's to increase portability
 - Tests underway with real LHC workloads
- CERN discussed CTA+EOS integration tests using Docker images and Kubernetes-managed cluster
 - Improves on past CASTOR integration tests
- Distributed computing @ IHEP
 - Leverage external resources to meet increasing requirements
 - Deploying services (VMDIRAC, JSUB, AWS, etc) to integrate all resources—more than doubled production from 2014 to 2016
- In coming years, growth in resource requirements by LHC experiments will exceed community's ability to provide them
 - Flat budgets & higher data rates from LHC are main causes
 - Find origins of bottlenecks and inefficiencies, seek alternative resources (ie, HPC, storage nodes) to cope

IT Facilities & Business Continuity

- Wigner described data center infrastructure
 - Upgrade of cooling for better control/regulation —tied to power consumption for higher efficiency
- CERN presentation on data center news
 - Water leak near electrical transformers—not yet fully resolved
 - Planning for new data center going forward
 - 2nd Network Hub—need for redundancy for critical networking equipment
- Orsay data center extension
 - Propose PUE < 1.3 with redundant power feed but no UPS—max load of 1.5 MW available to IT equipment
 - Cooling from rear-door heat exchanger—no CRAC's in data center
 - Estimated to be ready by Sept. 2018
- Managing hardware failure (almost automatically) at IN2P3
 - Cuts down time interfacing with vendor support staff
 - Use vendor (Dell in this case) API – not applicable to all failure modes
 - Very informative figures on historical failure data



HEPIX Fall 2017

- Next meeting at KEK Oct. 16-20
 - First HEPHX in Asia since 2012 meeting in Beijing
 - First HEPHX hosted by KEK
 - Co-located with HPSS Users Forum (HUF) and LHCOPN-LHCONE meeting

