
CERN

HEPiX Spring 2017 Summary

https://indico.cern.ch/event/595396

Hungarian Academy of Sciences

Jérôme Belleman • Jarka Schovancova • Luis Pigueiras

Outline

Jérôme

Spring 2017 Meeting & General HEPiX News

Computing & Batch Services

Basic IT Services

Site Reports

Jarka

Storage & Filesystems

Security & Networking

Grid, Cloud & Virtualisation

Luis

Miscellaneous

IT Facilities & Business Continuity

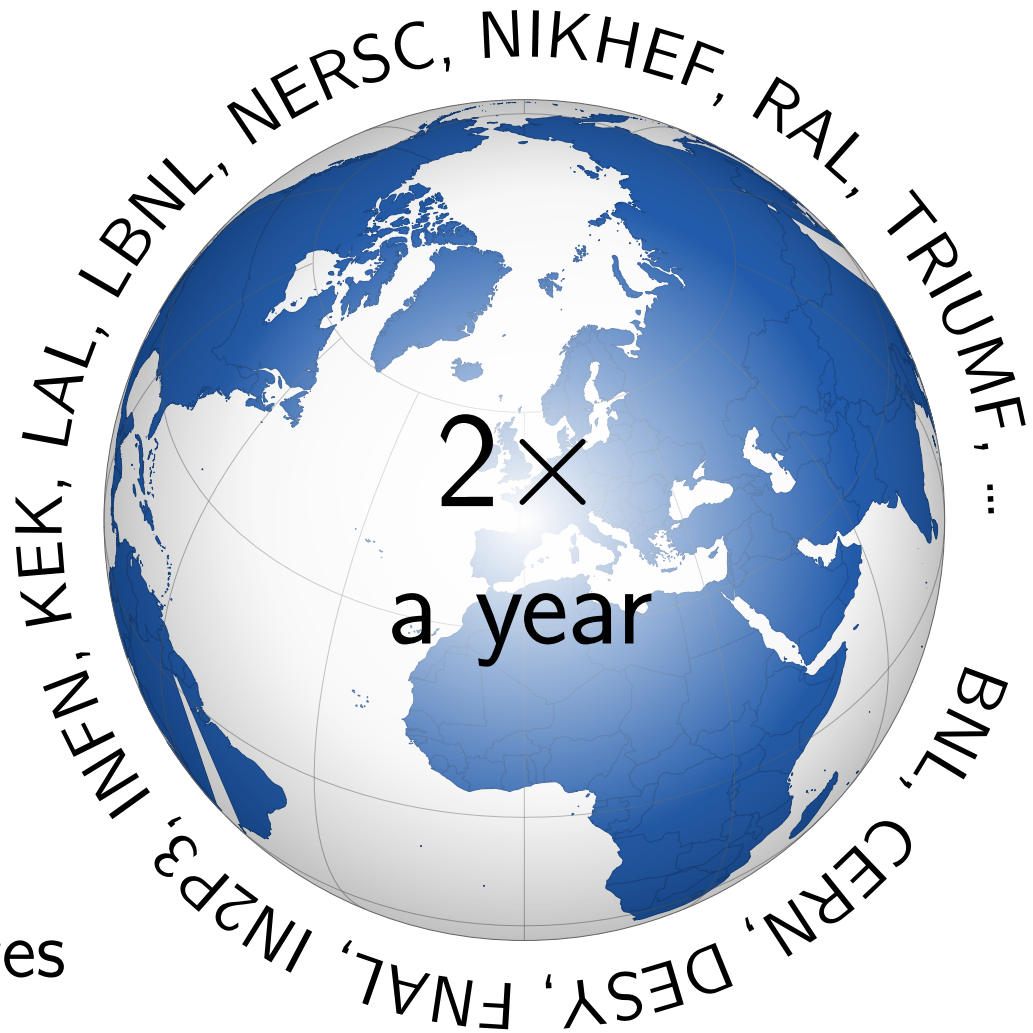
End-User IT Services & Operating Systems

HEPiX

Future plans

Working groups

Challenges



Recent work

Status reports

Experiences

http://www.hepix.org


upgrade
syslog ng
interface
Elasticsearch
Provider HEPiX
job
analysis
use case
resource
database
space
rack
group
user
solution
disk
HEPiX Budapest
standard
make
grid
Service Provider
OpenStack
scale
operation
many
system
use
test
experience
service
will
provide
message
machine
GB
Spring
CERN
easy
container
WLCG
running
different
total
status
Ceph
based
two
CTA
node
needed
https
tool
HS06
CERN
Security
model
instance
error
per
Main
cern.ch
local
VM
possible
HEPiX
Spring
issue
distributed
now
talk
set
high
fib
fib
link
RAL
Tier
FS
Service
Workshop
April
Salt
CernVM FS
current
setup
May
file
client
AFS
perfSONAR
capacity
Puppet
need
master
part
information
allow
Budapest
April
still
project
Spring Budapest
April Understanding
Linux
pilot
first
CVMFS
Spring Hard
well
small
testing
EOS
XRootD
PB
development
usage
tape
workflow
CPU
using
core
experiment
server
performance
module
time
problem
performance
DB12
central
platform
IPv6
CM
run
support
control
monitoring
available
better
software
Global
CernVM
Currently
TB
ATLAS
configuration
next
data
Evolution
Global
Thank
via
one
single
hardware
storage
log
API
check
level
Hard
Puppexit
deployment
Workshop
Budapest
day
also
started
new
production
power
version
Understanding
performance
Tier
Evolution
Spring
CCIN2P3
host
Singularity
year
since
done
plugin
large
dual stack
IHEP
MISP
change
CERN Cloud
team
etc
memory
used
network
Plan
Computing
cluster
site
Site Report
access
infrastructure
Spring Workshop
work
event
source
lib64
libpython2
Ivy Bridge
number
libpython2
lib64
application
cloud
several
dashboard
DB12
DB

The Hungarian Academy of Sciences

- First HEPiX in Budapest
- 125 participants
- Newcomers and long-time acquaintances
  SCS, JSI, Aachen, Zurich, KISTI, Tata, Alba, NRC
- Dinner cruise on Danube

The Wigner Data Centre

- 4 computer rooms, 3 for CERN
- 1 CERN room $1/2$ full, room for expansion
- Cool aisles for servers, hot aisles for UPS
- Target PUE < 1.5
- Optical room, nitrogen gas room
- Can sustain a 1000°C fire for 1 hour

Site Reports

- Belle II detector to collision point in SuperKEKB
- CERNBox, IHEPBox, BNL Box, ...
- PIC's oil-immersed servers
- *Refondation Project*: merge of HEP, astrophysics, accelerator laboratories
- DESY's *Science Data Management Centre*

https://indico.cern.ch/event/595396/contributions/2544084

Computing & Batch Services (I)

SLURM
High-Performance
Computing:

- PySlurm mentioned a few times
- Swiss National Supercomputing Centre working on fairshare

HTCondor
High-Throughput
Computing:

- Scheduling policies
- Group enforcements

Computing & Batch Services (II)

Intel's Knights Landing
(Wikipedia, The Next Platform):

- Many-cores processors
- Prefetching, pipe-lining, ... if coded right
- BNL, 9k-core cluster: RHEL, SLURM, Puppet
- JLab, 16k-core cluster: #397 on TOP500
- Runs HEP-SPEC06

Computing & Batch Services (III)

Benchmarking:

- Next generation?
- To compare with real workload
- 5 fast benchmarks
- Dirac Benchmark 2012, ATLAS Kit Validation
- Different implementations: C++, NumPy
- DB12 on Ivy Bridge and Haswell

Computing & Batch Services (IV)

CosmoHub:

- Galaxy analytics
- PostgreSQL $\rightarrow$ Hadoop
- 100$\times$ faster

Basic IT Services (I)

Miscellaneous monitoring technologies:

- Elasticsearch at CERN and elsewhere
- Sensu with RabbitMQ, Redis, InfluxDB
  - Schedules tests (e.g. with Nagios scripts)
  - REST API
  - Dashboards
- netdata is worth a look
- HEPIX Monitoring Special Interest Group

Basic IT Services (II)

Monitoring, a closer look at syslog-ng:

- CC-IN2P3 love it
- Central logging
- Reads from files, sockets, pipe, ...
- Classifies, structures, filters, enriches logs
- Pattern DB parses logs with XML descriptions
- To files, NoSQL DBs, messaging systems, ...

Basic IT Services (III)

Configuration Management:

- Other sites to Puppet 4
- SaltStack is very flexible
  - Push-based
  - YAML and Jinja2, Python-based
  - Event-driven, that's what it excels at
  - Steep learning curve

Security and networking

- Security
  - Security Operations Center & collaboration
- Network
  - IPv6 deployment
  - Network monitoring, performance

Security and networking (II)

Building and operating a large scale Security Operations Center (L. Valsan)

- https://indico.cern.ch/event/595396/contributions/2532562/
- Centralized system for the detection, containment and remediation of IT threats
- Ensures that security incidents are properly identified, analysed, reported, acted upon

Security Workshop (L. Valsan)

https://indico.cern.ch/event/595396/contributions/2566371/

- Hands-on session to install, configure, and operate MISP (Malware Information Sharing Platform & Threat Sharing)

Security and networking (III)

Computer Security Update (L. Valsan)
https://indico.cern.ch/event/595396/contributions/2532561/

- Presented updates in the past year

KEK Computer security update (T. Murakami)
https://indico.cern.ch/event/595396/contributions/2544140/

- Security infrastructure, fight against incidents
- Cooperative activities: exchange experience with other parties

Security and networking (IV)

Deployment of IPv6-only CPU on WLCG
- an update from the HEPiX IPv6 Working Group (A. Sciaba)
https://indico.cern.ch/event/595396/contributions/2548526/

- Much improved engagement by Tier-1s
  - Limited dual-stack storage at Tier-1s, but coming soon everywhere
- Good number of Tier-2s run dual-stack (but still minority)
- WLCG Tier 2s must start planning **NOW**
- Very few issues left, none critical
- Automatic endpoint monitoring in place

Security and networking (V)

WLCG/OSG Networking Update (S. Mc Kee)
https://indico.cern.ch/event/595396/contributions/2549875/

- Working infrastructure in place to monitor and measure our networks
- Report on perfSONAR updates
- New applications development in progress:
  - Notifications/alerting
  - Predictive capabilities
  - Current utilization and capacity planning
  - Evaluating network performance of commercial clouds

ESnet Update (J. Metzger)
https://indico.cern.ch/event/595396/contributions/2532572/

- Evolution in LHC community will impact network providers
- Changes in the ESnet and the other NRENs will impact LHC community
- Room for improvements in the LHC workflow network efficiency

Security and networking (VI)

IPv6 at the RAL Tier 1 (J. Adams)
https://indico.cern.ch/event/595396/contributions/2558578/

- RAL network ready to support IPv6 services
  - still some work to do, though
- Tier-1 on IPv6 in production
- Ready to go full dual-stack soon

Network related updates in IHEP (S. Zeng)
https://indico.cern.ch/event/595396/contributions/2557937/

- perfSonar deployed
- IPv6 testbed being prepared
- Network architecture will be updated in Summer
  - A distributed SDN architecture under R&D

Grids, Clouds, Virtualization

- Understanding and improving performance
- Resources information systems
- Operating clouds & containers

Grids, Clouds, Virtualization (II)

Understanding performance: optimisation activities in WLCG
(A. Kiryanov, A. Sciaba)
https://indico.cern.ch/event/595396/contributions/2532584/

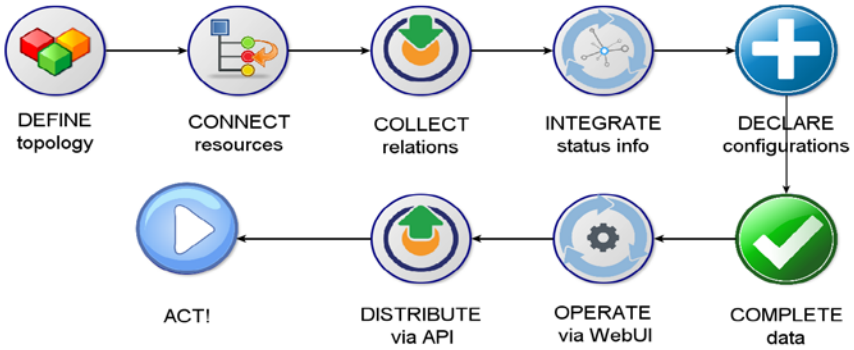
- Run jobs on storage nodes
- Storage and Computing can run without much interference
  - Typical I/O loads in production: can use >80% of the CPU on storage node for non-storage tasks

- WLCG: Stronger focus on increasing the usage efficiency
  - Fully exploit and optimise existing resources
- Contribute with tools and expertise
- Understand behavior of different workflows
- Prototype new ideas and concepts

Grids, Clouds, Virtualization (III)

The Computing Resource Information Catalog (A. Di Girolamo)
https://indico.cern.ch/event/595396/contributions/2532542/

- CRIC base platform of WLCG Information configuration system:
  - Building on top of successful experience with the ATLAS Grid Information System (AGIS)
- Architecture and core functionalities
  - Focus to needs of collaborations/experiments
  - Common framework for the description of all WLCG resources
  - Well defined & consistent interfaces



Grids, Clouds, Virtualization (IV)

CERN Cloud service update: Containers, migrations, upgrades, etc. (L. Pigueiras)
https://indico.cern.ch/event/595396/contributions/2544105/

- **Growing!** … and further expansions planned
  - Adding 2k compute nodes with ~86k cores
  - Adding new components such as Magnum and Manila
- **Facing major operational challenges**
  - Scaling with HW additions
  - Replacement of network layer

Grids, Clouds, Virtualization (V)

System testing service developments using Docker and Kubernetes: EOS + CTA use case (J. Leduc)

https://indico.cern.ch/event/595396/contributions/2532576/

- Reproducible development environment, allows regression tests
- Fast, flexible, isolated and self contained in software repository

Container Orchestration - Simplifying Use of Public Clouds (I. Collier)

https://indico.cern.ch/event/595396/contributions/2556631/

- Kubernetes: a promising way of running LHC workload
- Portability between on-premises resources & public clouds

Grids, Clouds, Virtualization (VI)

ElastiCluster - automated deployment and scaling of computing and storage clusters on IaaS cloud infrastructures (R. Murri)

https://indico.cern.ch/event/595396/contributions/2548531/

- CLI tool to create, set up and scale computing clusters hosted on IaaS cloud infrastructures
  - Configuration: Ansible
  - Supported computational clusters: batch-queuing systems, Spark/Hadoop, Mesos + Marathon
  - Distributed storage: GlusterFS, HDFS, OrangeFS/PVFS, Ceph
  - Interfaces to multiple cloud flavours: OpenStack, AWS EC2, GCE
  - OS images: CentOS 6/7, SL6, Debian 7/8, Ubuntu 14.04/16.04

Distributed computing in IHEP (X. Zhang)

https://indico.cern.ch/event/595396/contributions/2557936/

- General computing platform supporting multiple experiments
  - Re-using common existing tools to serve wide range of communities

Storage, filesystems, data mgmt

- Strategy planning
- Storage for collaborations
- Ceph
- Advances in storage technologies

Storage, filesystems, data mgmt (II)

CERN IT-Storage Strategy Outlook (L. Mascetti, J. Leduc)
https://indico.cern.ch/event/595396/contributions/2557938/

- Vision for CERN Storage:
  - Build a flexible and uniform infrastructure
  - Maintain and enhance (grid) data management tools
- EOS, CERNbox, Data archiving, CTA (CERN Tape Archive)
  - Streamline data paths, software and infrastructure

Storage, filesystems, data mgmt (III)

EOS and CERNBox Update (L. Mascetti)
https://indico.cern.ch/event/595396/contributions/2544082/

- EOS provides very flexible platform for large communities
- CERNBox: on top of EOS, extension of Desktop, provides storage for 9k users
- Strategic direction for CERN based disk storage
  - for physics data (user/group/grid)
  - as generic home directory via EOS fuse and CERNBox

BNL Box (H. Ito)
https://indico.cern.ch/event/595396/contributions/2556634/

- Built on top of OwnCloud (as CERNBox is)
- Users can store, access and share their data anywhere

Storage, filesystems, data mgmt (IV)

Federated data storage system prototype for LHC experiments and data intensive science (A. Kirianov)
https://indico.cern.ch/event/595396/contributions/2532587/

- Working prototype of federated storage
  - 7 Russian WLCG sites organized as 1 homogeneous storage with a single entry point
- Validation of infrastructure
- Technologies: Exploited EOS and dCache; Plans with SW solutions (HTTP-based, DynaFed)

RAL Tier-1 Evolution as a Global CernVM-FS Service Provider (C. Condurache)
https://indico.cern.ch/event/595396/contributions/2532590/

- Topology:
  - Cvmfs stratum-0 for EGI and STFC
  - Cvmfs stratum-1 for WLCG, EGI, and STFC
  - Cvmfs uploader service (EGI, STFC)
- Developments for “protected” cvmfs repositories with X509 proxy
- Working prototype at RAL; also tested in cloud

Storage, filesystems, data mgmt (V)

An update to Ceph at RAL (T. Byrne)
https://indico.cern.ch/event/595396/contributions/2553417/

- ECHO in production
  - Erasure Coded - Ceph - High throughput - Object store
- Support for GridFTP/XRootD on a Ceph object store are mature
- Plans to support industry-standard protocols

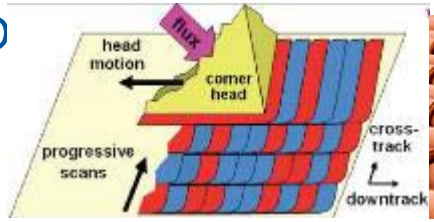
Data-NG: A distributed Ceph infrastructure (G. PHILIPPON)
https://indico.cern.ch/event/595396/contributions/2558271/

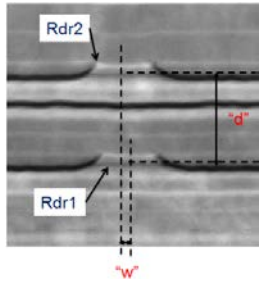
- Building a storage infrastructure to support P2IO labs
- Aim for resiliency and efficiency; 100 GB/s among sites
- Future extension with cost proportional to added capacity

Storage, filesystems, data mgmt (VI)

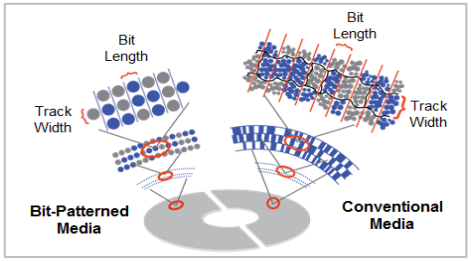
Advances in storage technologies (J. Fagan [Seagate])
https://indico.cern.ch/event/595396/contributions/2558799/

- Challenges towards 100 TB HDD
- Technology for high capacity
  - Shingled Magnetic Recording
  - Sealed Drive Technology





- High capacity/performance: 2 or more readers on 2 adjacent tracks
- Technology leadership (capacity)
  - Heat-Assisted Magnetic Recording (HAMR)
  - Heated Dot Magnetic Recording



- Need for more IOPs/TB
  - Higher capacity drives (>10TB) create IOPs/TB constraints, many customers require 8 - 12 IOPs/TB
  - With new HDD recording tech (e.g. HAMR) the IOPs/TB drops below acceptable levels without other tech mitigations
- Overview of Storage Utilization in the Data Center and Cloud
  - Mix SSD and HDD by application

IT Facilities and Business Continuity (I)

- 4 presentations
- Wigner presented a description about their DC
  - Update in cooling problem
  - Solved with new regulation system
- CERN (by Wayne) presented some issues + updates
  - Water leak on electrical transformers and power cut (9 March)
  - Current state of the 2nd Datacenter in Prevessin and 2nd Network HUB

IT Facilities and Business Continuity (II)

- P2IO/LAL with a DC extension in Orsay
  - Been in prod since 2013
  - Increase from 30 to 50 racks
  - Prod estimation for September 2018
- Manage your hardware failures in an (almost) automatic way (by IN2P3)
  - Hardware detection problems with custom code called Nono
  - Communication with Redmine and Vendor API to handle solution

Miscellaneous

- Proposal for a journal dedicated to "Software and Computing for Data-intensive Physics"
  - From Spring HEPiX 2016
- Launched this winter a journal called *Computing for Big Science Journal*
  - Infrastructure for large-scale
  - High-throughput computing
  - Deep learning
  - Distributed data analysis
  - ... and anything related with Big Science!
- Looking for people to collaborate

End-User IT Services & Operating Systems

- Only one presentation about this (and it was CERN :()
- We presented:
  - Current OSes we are using (SLC5 with EOL, SLC6, RHEL5/6/7 and CC7)
  - Lifecycle of CC7 releases (from C7 release $\rightarrow$ testing $\rightarrow$ CC7 release)
  - And some numbers! (like proportion between OSes and Koji builds)

Next Meetings

KEK
Tsukuba, Japan
16 to 20 October 2017

University of Wisconsin
Madison, USA
14 to 18 May 2018

http://www.hepix.org

Questions?

www.cern.ch