

CCIN2P3 Site Report

HEPiX Spring 2015 Workshop

Mar 24th 2015 / Julien Carpentier
On behalf CCIN2P3 Staff

julien.carpentier@cc.in2p3.fr

- **BATCH / WLCG**
- **Evolution of the Computing Power**
- **Storage Summary**
- **Systems Administration**
 - Cloud services
 - Systems Management
 - Security & Authentication
 - Monitoring
- **Hosted Projects (eTRIKS)**

BATCH Running version UGE 8.1

Model	Cores	Nodes	Cores total	HS06 / Cores	Total HS06	Power %
C6220 II	40	141	5840	11,35	64 014	28,12
C6220	32	303	11232	9,67	93 760	41,17
C6100	24	297	7128	8,62	61 443	26,98
M610	16	62	992	8,57	8 501	3,73
Total		856	23 456		227 719	

Next Migration to version UGE 8.2

Status : still in evaluation

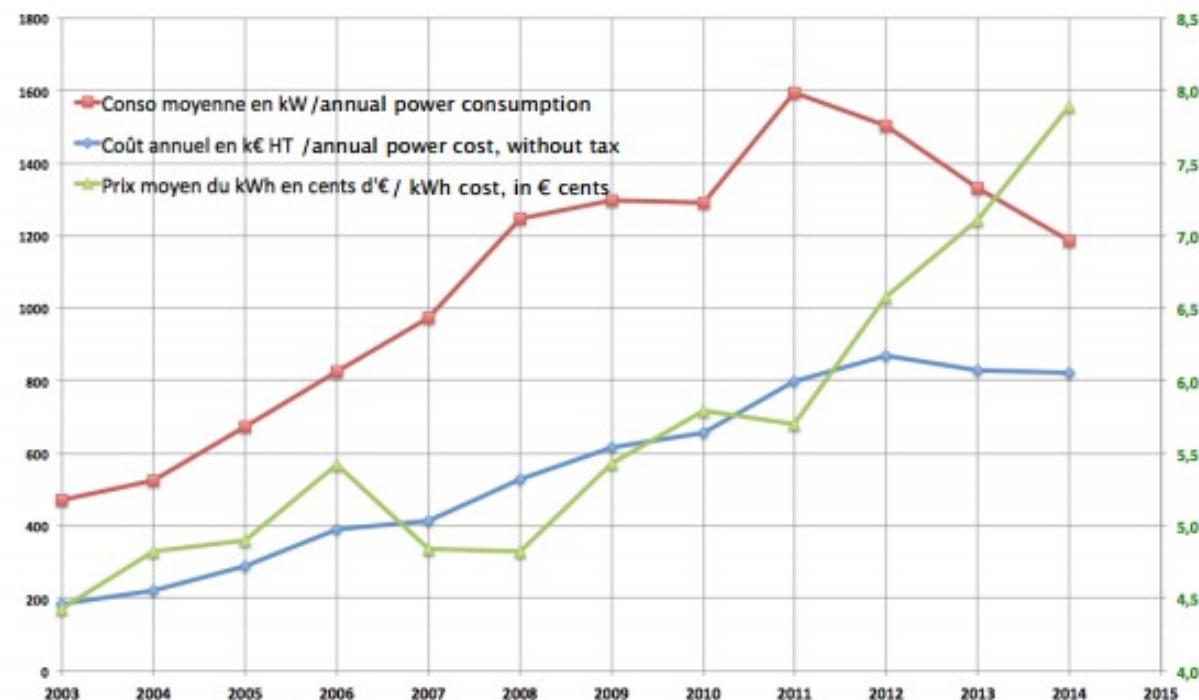
Key features : cgroups, multi-thread

WLCG : Tier-1 and Tier-2

LHC pledges (T1+T2)		2014	2015	Growth	
	CPU	101 kHS	125 kHS	+ 24 kHS	+ 24 %
	Disk	9.7 PB	11.6 PB	+ 1.9 PB	+ 20 %
	Tape	11.5 PB	18.1 PB	+ 6.6 PB	+ 57 %

Evolution of the computing power

- ▶ 5% to 7% of annual power cost increase
- ▶ Looking for IT efficiency
- ▶ Fine tuned power contract helps to minimize costs



Seek the more efficient IT hardware
in the most efficient room

Storage Summary

■ HPSS (updated 2015-03)

- 25 PB, 47 M files on tapes / 480 TB on disks
- HPSS V7.4.1.2e2 on RHEL 6
- Core server: DELL R720
- Metadata (DB2) : MD3220 SSD / SAS
- Tape drives: 60 x T10K-B / 22 x T10K-C / 20 x T10K-D
- 15 x tape servers
 - 5 x IBM AIX (p520) for T10K-B
 - 10 x Dell PowerEdge R710/720 RHEL6
- 14 x Dell disk servers
 - PowerEdge R510 + PowerVault
 - MD1200 + 10Gbps Eth + RHEL6
 - Capacity : 36 TiB / 54 TiB
- ACSLS 8.3

■ iRods

- V3.3.1
- 15 servers, 804 TiB
- Over than 9PB, ~92 M files

■ xrootd

- V3.0.5, v3.3.6, v4.1
- 44 servers R510/R720 (SL6)
- 2.5 PiB

In progress evaluation :

- Ceph
- Swift improvement (data life-cycle)

■ OpenAFS

- v1.6.1 and v1.6.2 (servers), v1.6.1 (worker nodes)
- 42 GiB
- 42 servers, ~1400 clients

■ SPS (GPFS) (updated 2015-03)

- V4.1.0
- 500 M files, 39 filesystems (80 groups)
- 2.1 PiB (mainly on R620/630 + MD3260/3460)
- 1000 clients, 60 servers
- Soon metadata on Flash
 - NetApp EF560 with 12 TB

■ TSM (updated 2015-03)

- 4 servers (AIX 6, TSM 6) on DCS3500
- 1 billion+ files, ~1 PiB
- 20 LTO4 drives, ~2000 LTO4 tapes
- 6 LT06 on IBM TS3500
- 3 TiB / day

■ dCache (updated 2015-03)

- 2 instances: LCG (v2.10.21) & EGEE (v2.10.13)
- +300 pools on 165 servers
 - (Dell R510, R720/730 on SL6/7)
- ~8,5 PiB
- 60 M files, 7.5 M directories

Deployment :

- Scientific Linux 6 (requirement for 7 in Kilo release)
- Griddynamics, then EPEL and now RDO
- Configured with Puppet



Resources :

Bunch of C6xx, R6xx, M6xx DELL PowerEdge servers

Core services

- 150 CPUs
- 300GB RAM
- 5TB storage

Hosting

- 50 CPUs
- 200GB RAM
- 4TB storage

Computing

- 500 CPUs
- 3TB RAM
- 9TB storage

Preprod cluster :

- 300 CPUs
- 1TB RAM
- 28TB storage

+30 TB Cinder volumes

+24 TB Swift S3 storage

Grand total:

- 1k CPUs
- 4.5TB Memory
- 100TB storage



TESTING & PREPRODUCTION

For whom : CCIN2P3 people to provision their testing and development systems

Goals

- Self service
 - host lifecycle handled by end users
 - accurate sizing and environment specification
- Test systems are most of the time resources thrifty
→ efficient mutualization
- On top of it : puppet enables reproducibility for the whole, moving from preproduction to production is handy

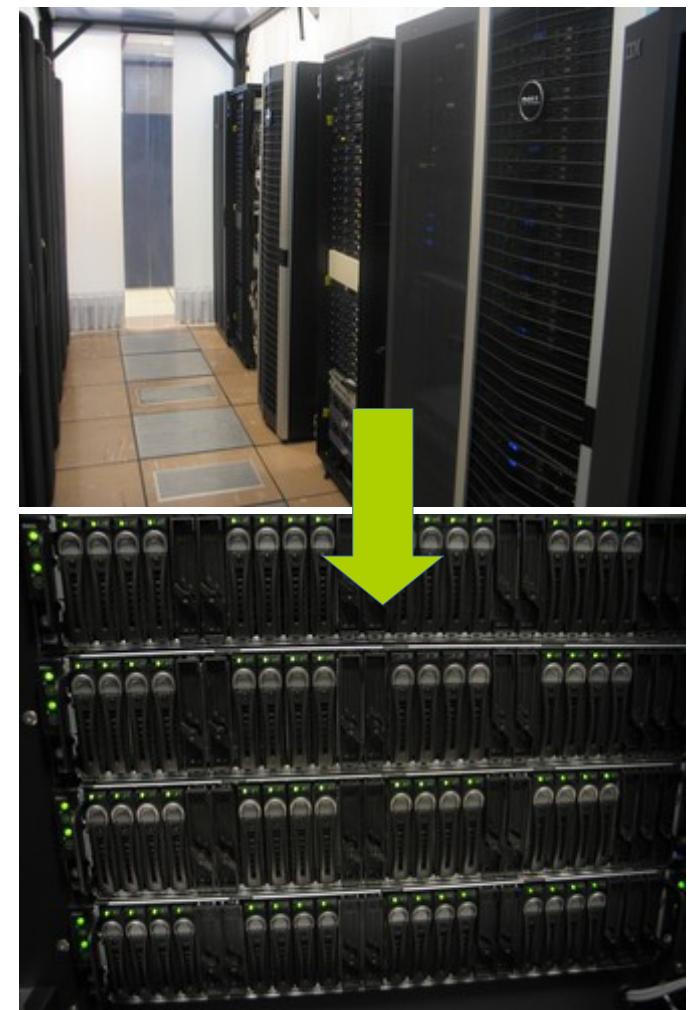
DEPLOYING CORE SERVICES

For whom : CCIN2P3 Sysadmin team provisionning core services

Goals

- Usual virtualization motivations
 - lowers hardware management
 - optimizes resources (mutualization)
 - soft servers sizing (cpu/ram/disk/net)
 - unbind servers from the hardware (higher availability, maintenance eased)

We formerly used VMware



COMPUTING

Users motivations :

- Gain access to opportunistic resources (fast access during activity bursts)
- Use specific environment (OS, softwares...)
- Implement their own scheduling
- Ease software deployment

Different models :

- The local batch system is aware of the cloud (statically, dynamically)
- Grid middleware leveraging cloud interfaces (Dirac...)
- Direct use of cloud interfaces (EC2/Nova...)

First productions since late '14:

- Atlas for MC jobs (opportunistic, simulation)

Currently evaluated cloud computing models :

- Large Synoptic Survey Telescop (<http://www.lsst.org>)
- Euclid (<http://www.euclid-ec.org/>)
- Bioaster

Cloud services (Roadmap)



Openstack components :

Operational :

Nova
Horizon
Cinder
Ceilometer
Swift

Still in evaluation :

Neutron



Incoming improvement :

High-Available DB Backend
with MariaDB & Galera cluster





Puppet at CC-IN2P3

Since 2009

Infrastructure :

- 3 servers
- Over than 1400 clients (Linux + AIX)
- About 150 https request per sec

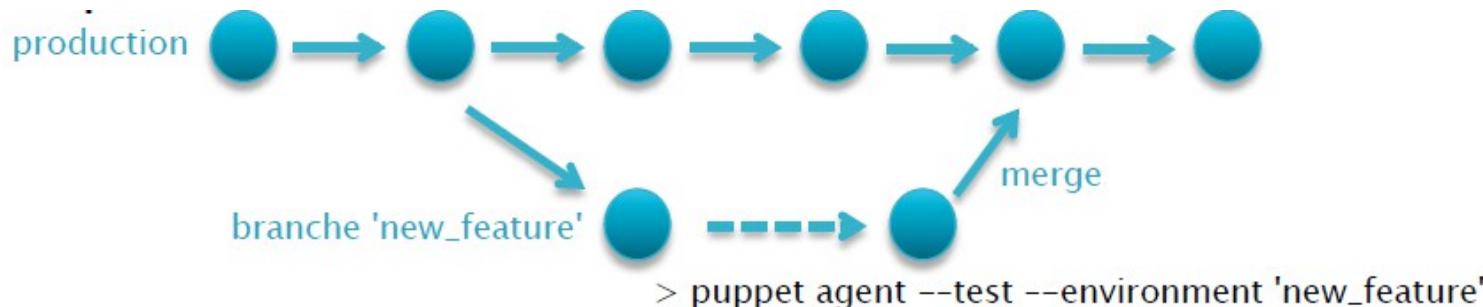
New features

Hiera and Environment with Gitolite

→ allow privileges delegation for service admins

Git/Gitlab backend with multiple branchs

- collaborative works
- functional tests (pre-production)





Replacement for our home-made installer BOA
Why ? BOA only for SL5/6, No support

- Open source, Cross-platform, Ruby & Javascript
- DHCP, TFTP, PXE for bare-metal, KVM, VMWare
- Dashboard or Command in-line (hammer-cli)
- PuppetDB Proxy

	Nom	Système d'exploitation	Environnement	Modèle	Groupe d'hôtes	Dernier rapport	
<input type="checkbox"/>	N ccosndl0025.in2p3.fr	Scientific...	production_git	PowerEdge R630	SL 6		<button>Éditer</button>
<input type="checkbox"/>	N ccosndl0026.in2p3.fr	Scientific...	production_git	PowerEdge R630	SL 6		<button>Éditer</button>
<input type="checkbox"/>	N ccosndl0027.in2p3.fr	Scientific...	production_git	PowerEdge R630	SL 6		<button>Éditer</button>
<input type="checkbox"/>	N ccosndl0028.in2p3.fr	Scientific...	production_git	PowerEdge R630	SL 6		<button>Éditer</button>
<input type="checkbox"/>	N ccosndl0029.in2p3.fr	Scientific...	production_git	PowerEdge R630	SL 6		<button>Éditer</button>
<input type="checkbox"/>	B ccsvli47.in2p3.fr	Scientific...	production_git	VMware Virt...	SL 6		<button>Éditer</button>
<input type="checkbox"/>	N cctest28.in2p3.fr	Scientific...	production_git	KVM	SL 6		<button>Éditer</button>



Keberos 5

Migration from KAS (Kerberos 4) used by AFS
Why ? Kerberos 4 only support Data Encryption Standard

Used for :

- Batch Grid Engine
- Machines Access
- Privileges delegation with remctl (remote control)
- Interactive Access (AFS)

Current limitation compare to AFS/Kerberos 4 :

- No centralization for authentication failure

→ Brute force protection with Fail2Ban

 Protection on interactive SSH

 Restrict number of attempts

 Reject blacklisted IPs (route)

Events

- logs : 2000 per seconds (40000 peaks)
- sources : 2000 servers
 - 300 Electrical Equipments
 - ~ 400.000 metrics

1 Billion events per day !

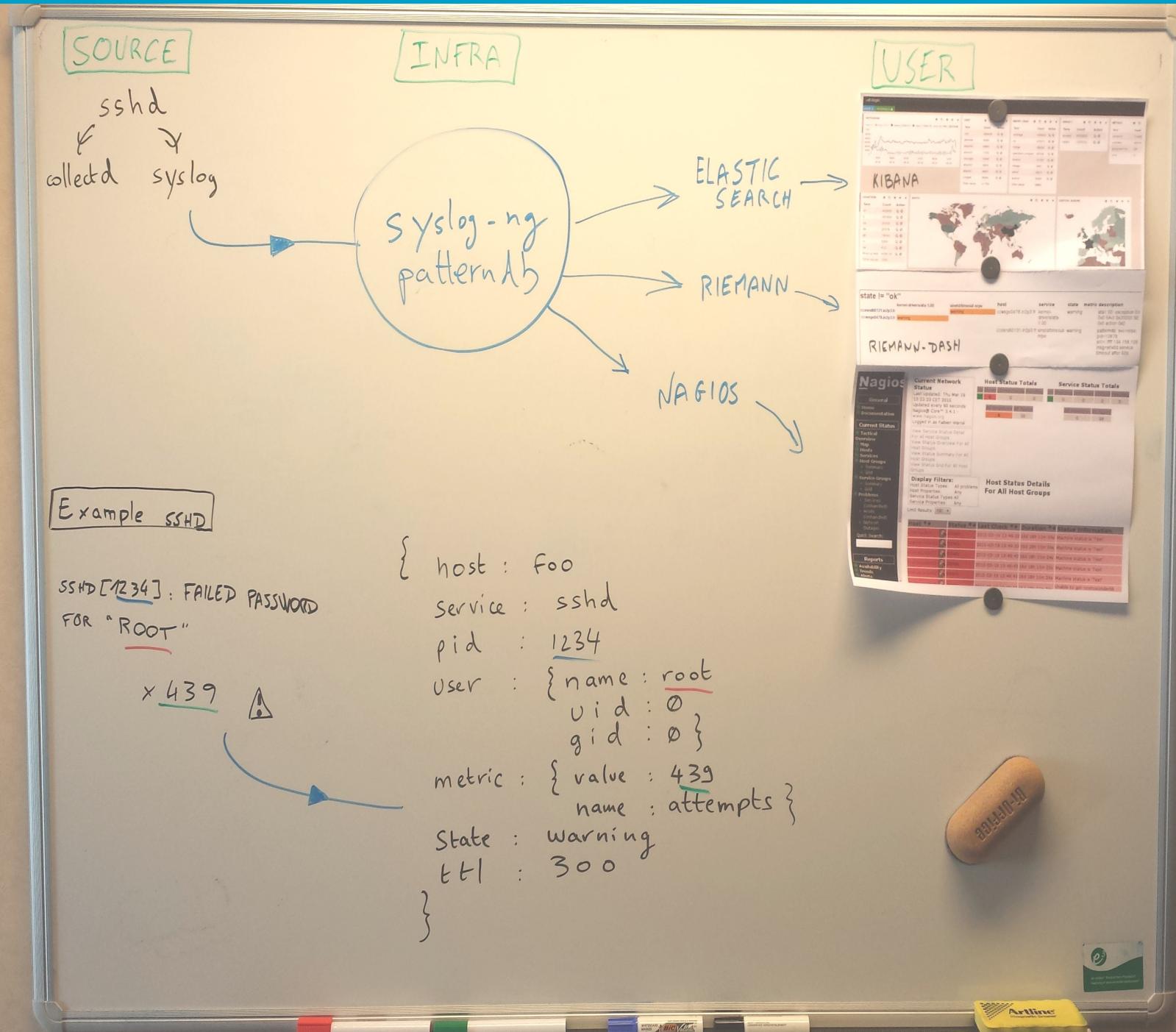
Tools used at CCIN2P3

- Metrics collector : Collectd
- Analyzer : Syslog-ng
- Real-time Engine : Riemann
- Storage and indexing : ElasticSearch
- Web-interface : Kibana, Riemann-dash
- Alerting : Nagios

Hardware

- ElasticSearch cluster : 6x R620 (48GB RAM, 3.5TB disk)
- Metric storage (collectd) : 4x R610 (Basic config)
- Other services : Few Virtual Machines

Monitoring (Architecture)



ETRIKS European Project



Budget

23.79M € for 5 years (Oct 2012 to Sept 2017)

Members

10 Pharma, 3 Academics, 1 standards, 2 Commercial suppliers

Goal :

- Provision of a KM Service to support Private / Public Translational Research
- Common platform and single access point for curation and computing

Hardware & Software :

- TranSMART : Java Web Application (KM Platform)
- Servers : Openstack dedicated nodes & DB servers
→ 208 vCPUs / 1.1 TB RAM / 130 TB Block Storage / 114 TB DB Storage
- R : Statistical computing

감사합니다 Natick
Grazie Danke Ευχαριστίες Dalu
Grazie Thank You Köszönöm
Grazie Спасибо Dank Tack
谢谢 Merci Gracias
谢谢 Merci Seé
谢谢 Merci ありがとう

Questions ?

