Monitoring multi-PB disk farms at CERN for the HEP experiments:

a service manager view

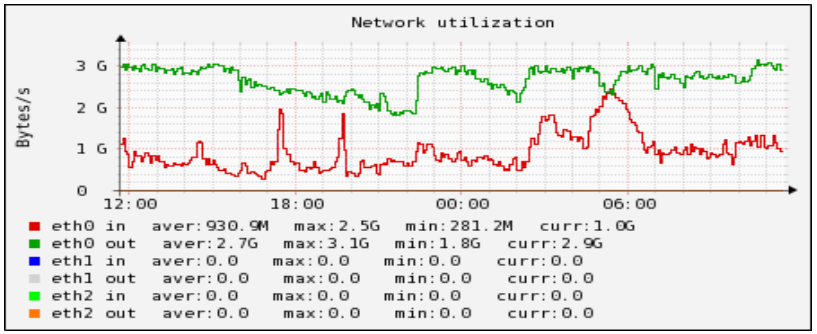
HEPiX Spring 2013 Workshop
15-19 April 2013
CNAF Bologna (Italy)

Massimo Lamanna / CERN
IT department

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

~ 3 years ago...

- Running a complex disk farm (CASTOR) with ~o(10^3) disk servers plus a number of “special nodes” (head nodes, SRM, ..)
- Online + historical time series (LEMON and SLS for example for CPU usage and available disk space)
  - Rich set of quantities with a good interface
  - Static set of quantities, limited interactivity



- ~30 GB/day of log (o(10)TB/year)
  - Complete and nicely formatted raw info
    - But “all-or-nothing” (need to be root on all machines, sensitive info)
    - “Just-in-case” usage (cron usage does not really scale)
  - But flexible, parallel and with some history functionality (log rotation)

See next page →

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

wassh/awk/grep/uniq/sort hell (or heaven)


```
wassh -q -l root -c c2public/server 'cat
/var/log/castor/stagerd.log.1 | grep
StagePrepareToGetRequest' | grep SvcClass=\"amsuser\" | sort
| cut -c 15-16 | uniq -c | awk '{print $1 " events per hour
at " $2":00 h"}'
45995 events per ho    at 04:00 h
61754 events per h     at 05:00 h
115032 events per      at 06:00 h
153544 events per      at 07:0
103910 events pe       at 08:
.
```

Esoteric
Cannot share it with anybody!

Powerful: frequency of the action of a subset of users summed up on all headnodes a yesterday

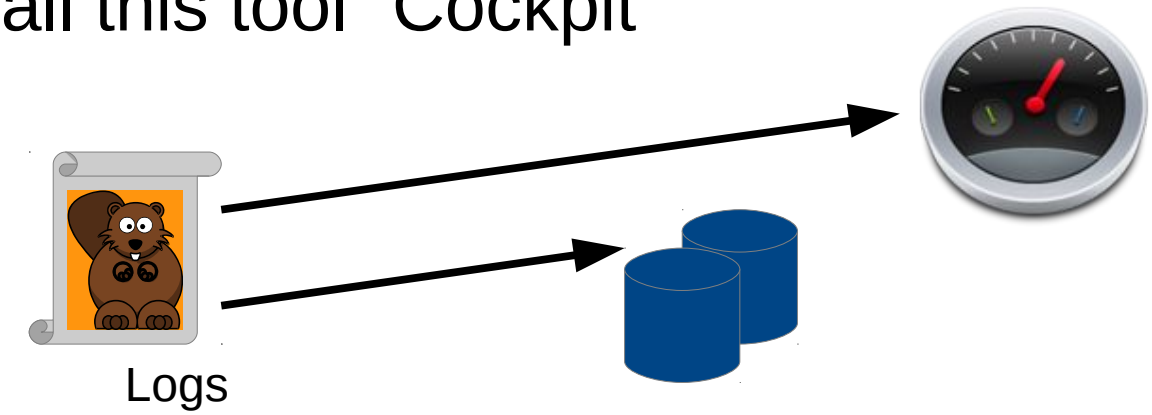
Not any good for “Empowering the users!”…

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Starting points (simplified assumptions)

- All “actions” (e.g. file creation) in logs
  - 30–60 GB/day
  - Distributed across o(10^3) nodes
- All “load indicators” in Lemon or SLS
  - e.g. I/O of a (set of) boxes or number of scheduled transfers
- Make available all quantities in a semi-interactive way with a flexible display
  - and save all logs in a long-term repository

… we call this tool “Cockpit”



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

DSS

Multiple facets

CERN IT Department

- **Correlation engine**
  - Behaviour of a given box/daemon (vs time)
    - e.g. SRM request rate
  - Build the “trajectory” of a file in the system
    - e.g. File created, registered, received, read, migrated to tape, garbage collected, recalled...
  - Correlation with other components of the computing infrastructure of the experiments
  - e.g. Stager request per second vs CPU usage of the same boxes
  - Correlate events and system components
    - e.g. Evaluate probability for disk failures on a given service class of hardware type
- **Display**
- **Log repository**

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CERN

Multiple facets

- **Correlation engine**
- **Display**
  - Present our system behaviour (also to users, managers, …) in a straightforward way
  - Combine high-level information using additional sources
    - e.g. Detection of misconfigured nodes directly published (no log)
- **Log repository**

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Multiple facets

- **Correlation engine**
- **Display**
- **Log repository**
  - Disaster recovery might need access to very old information
    - Normally we keep log files in place for several months
    - Main use case: recover file metadata for files removed from the disk and the name space/catalogue
  - Similar requirements for other auditing

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Cockpit status

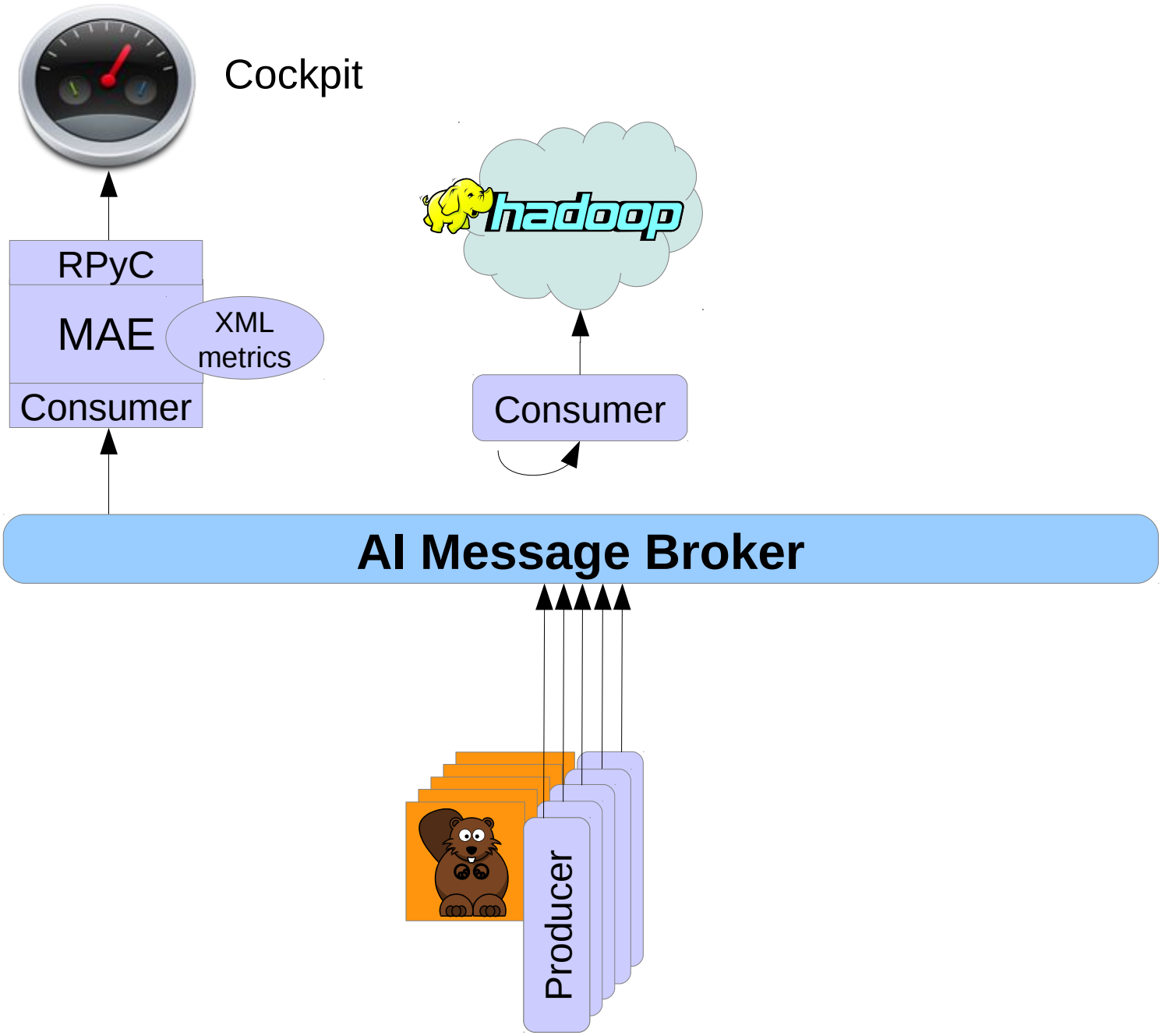
- **Highlights of the system being built**
  - Our initial system being blended in the Agile Infrastructure monitoring
  - Conceptually very similar to our initial prototype but augmented by AI
  - Collaborative spirit
- **Examples:**
  - **Replace/integrate the existing monitoring**
  - **What about `wassh|cat|grep|awk` gymnastic?**
  - **Interactive access to monitoring data!**
  - **Publish KPIs**
    - Used by power users as well (migration backlogs, disk-cache lifetime)
  - **Debugging**
    - **Day-by-day activity in operations**
  - **Client migration**
  - **Hardware inventories**
  - **Alarms**
  - **Log repository**

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

DSS

Zero-level approximation

CERN IT Department



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CERN

How we use it

Cockpit

Lemon sensors

LogViewer (Visualisation)

RPyC

MAE

CASTOR metrics

Consumer

hadoop

Consumer

APACHE HBASE

AI Message Broker

Lemon

Producer

Status of machines

Nameserver logs

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Metrics analysis engine

CERN IT Department

- XML-like terse and powerful syntax
- Create a file, cp in the right directory, open the browser
  - Just-in-time. Drop the metric description at time t0
  - Once uploaded, the metric starts being filled and it is visible in the browser
  - Data for t>t0 will be available (Cockpit mysql DB)

```
<metric>
   name: CountLogProcessingFileQuery
   UnitL Nb of Log
   category: General
   window: 60
   conditions: LVL in ["Info"] and MSG =="Processing File Query
   groupbykeys: INSTANCE
   data: Counter(COUNT)
   handle_unordered: time_threshold
   nbins: 1
</metric>
```

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CERN

Another example

```
<metric>
name: Errors
unit: Count of errors
category: General
window: 60
conditions: LVL in ["Error"]
groupbykeys: INSTANCE, DAEMON
data: Counter(COUNT)
handle_unordered: time_threshold
nbins: 1
</metric>
```

Output :

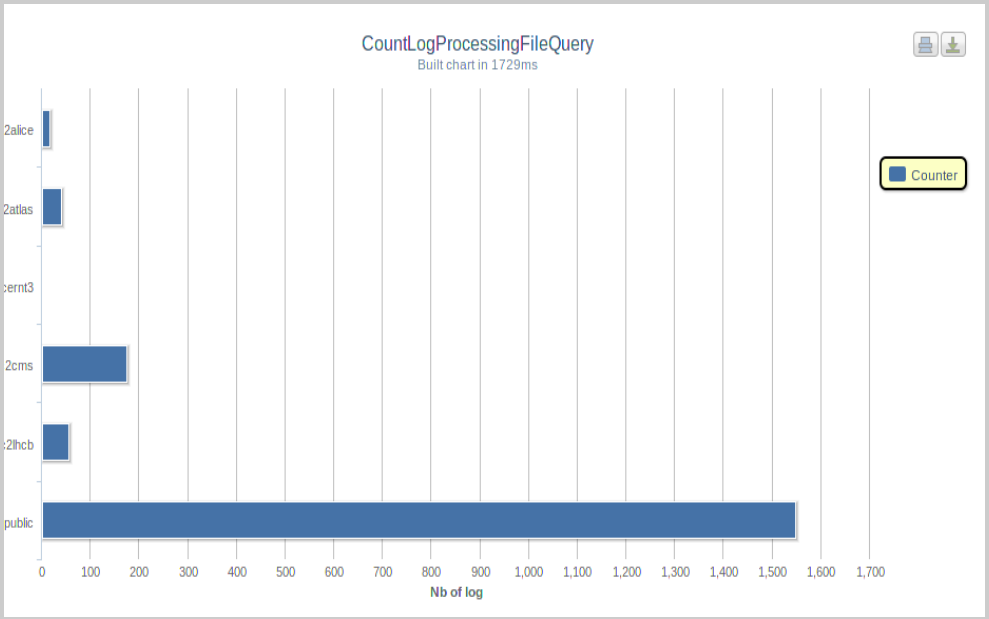
- c2alice
  - nsd : 12
  - stagerd : 14
  - transfermanagerd : 8
- c2atlas
  - nsd : 4
  - stagerd : 120
  - transfermanagerd : 28
- c2cms

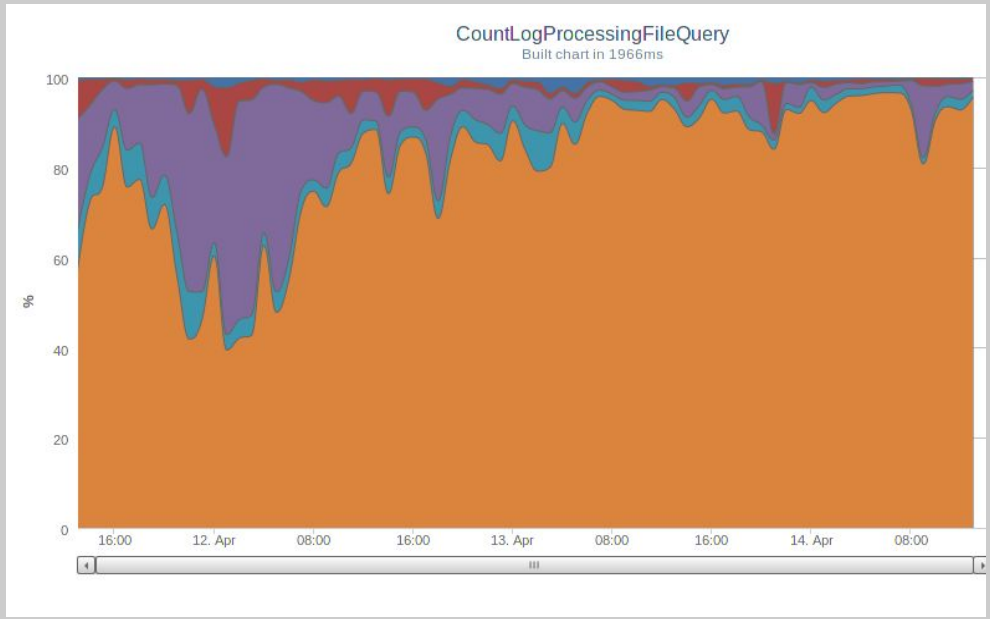
...

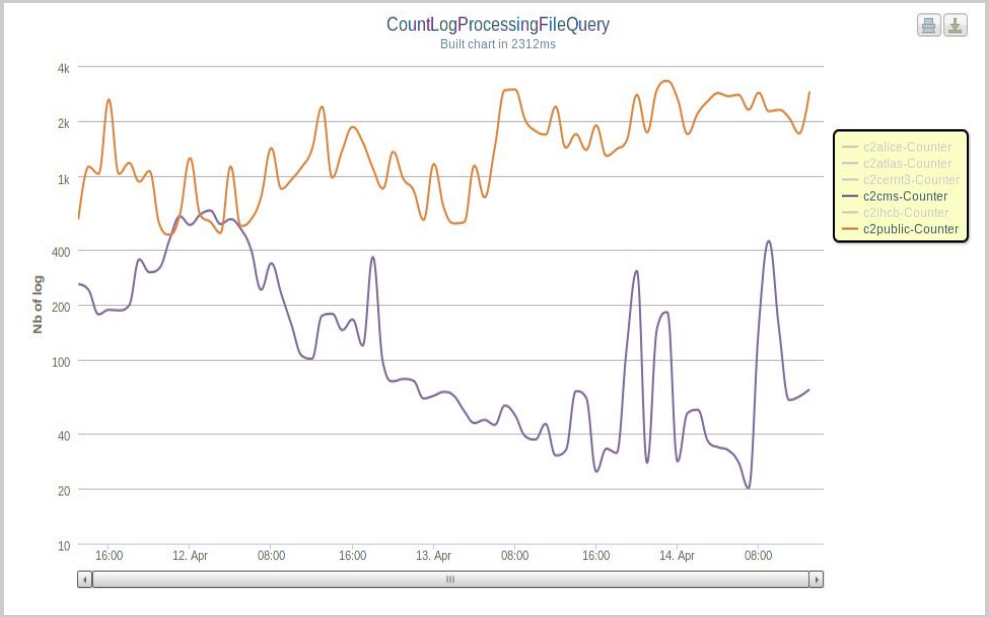
- t<t0?
  - For historical data one can Hadoop/MapReduce stored logs.
    - Offline analysis
    - Feed data back to Cockpit DB (prototype level)

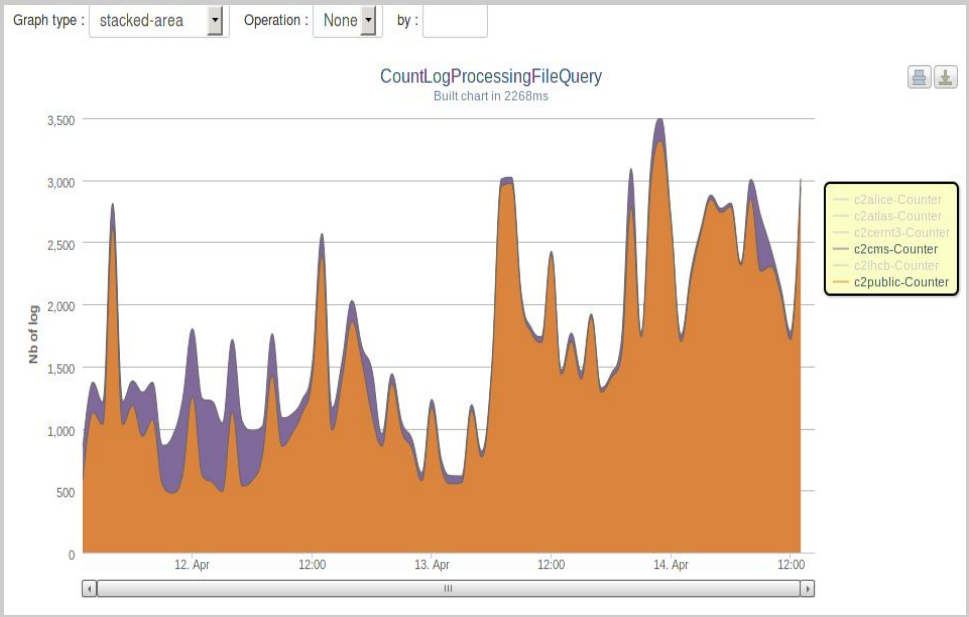
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Cockpit examples









(Min) Lifetime of 1 pool (2013 Q1)

Lifetime in weeks

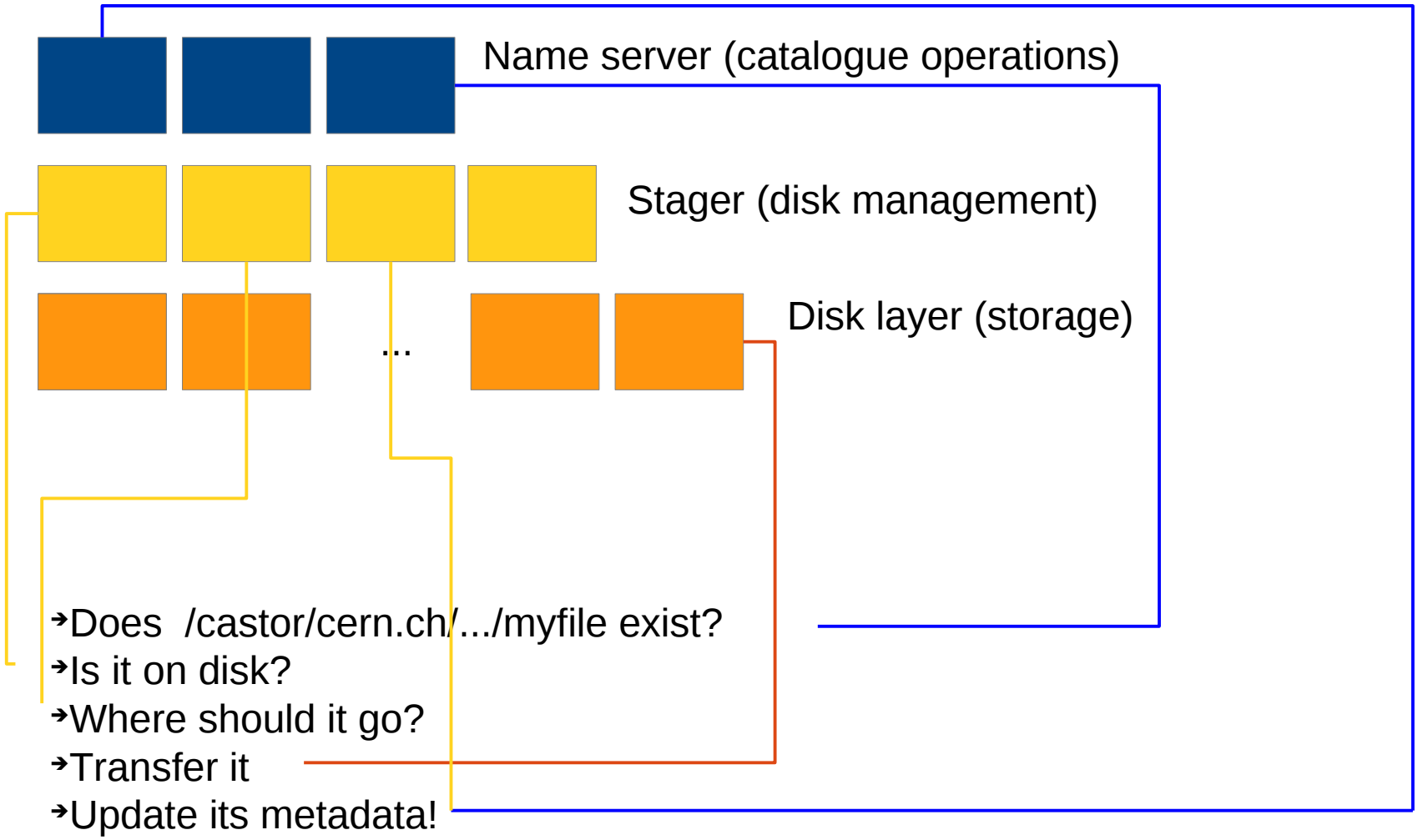
From 1/1/2013 to 31/1/2013

- Click for: JPEG/PNG/PDF/SVG
- CSV?: http://c2adm01/metric/GCFileAgePublic + curl + simplejson get the raw data and play with your favorite tool

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Actions in action

e.g. write a file: **`rfcp myfile /castor/cern.ch/.../myfile`**



- *Each client action is a sequence of actions on distributed entities*
- *The life of a file is a sequence of synchronous and asynchronous actions*
  - *client actions like the creation*
  - *Internal actions like tape migration*

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

DSS Log viewer

CERN IT Department

Replacement of its predecessor (based on ORACLE)
Underlining technology: Hadoop/HBase
In production for 6+ months (DLF switched off and discontinued)

File ID : NSFILEID Request ID : REQID Tape ID : TPVID Search Reset

Query : **File ID == 1207778773**

Show 10 entries Show / hide columns Search columns Treat as regexp : String or Regexp Search Reset

Showing 1 to 10 of 95 entries

First Previous 1 2 3 4 5 Next Last

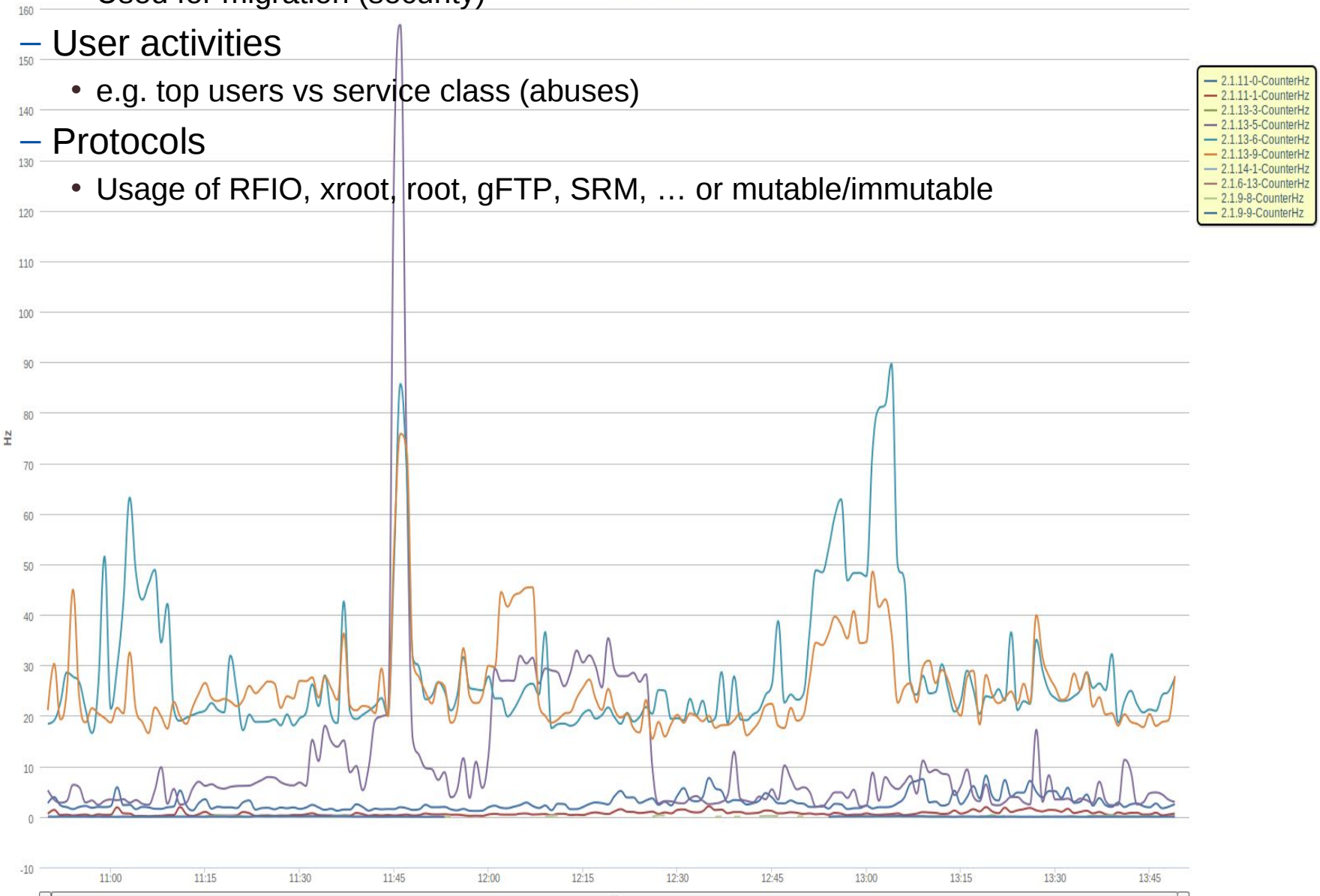
Timestamp	Severity	Instance : Hostname	Daemon	PID	TID	Message text	Request ID	Tape ID	Payload
2013-01-08 16:43:58.723374	Info	c2alice : lxfsrc1108	gcd	1479	1482	Removed file successfully	-	-	**GcType**=Too many replicas **NbAccesses**=0 **SvcClass**=t0alice **GcWeight**=1355815613.461748 **NSHOSTNAME**=castorns **Filename**=/srv/castor/05/73/1207778773@castorns.14993392337 **FileSize**=532521861 **LastAccessTime**=1355815494 **FileAge**=1844225
2013-01-08 16:43:58.685670	Info	c2alice : c2alicesrv301	stagerd	22093	22101	File selected for deletion	35651540-4d19-4f93-9b7a-00274783d906	-	**Filename**=/srv/castor/05/73/1207778773@castorns.14993392337 **NSHOSTNAME**=castorns **DiskServer**=lxfsrc1108.cern.ch
2013-01-08 16:33:45.241274	Info	c2alice : lxfsrc1107	gcd	3529	3531	Removed file successfully	-	-	**GcType**=Too many replicas **NbAccesses**=0 **SvcClass**=t0alice **GcWeight**=1355944313.515305 **NSHOSTNAME**=castorns **Filename**=/srv/castor/03/73/1207778773@castorns.14999559347 **FileSize**=532521861 **LastAccessTime**=1355944314 **FileAge**=1714921
2013-01-08 16:33:45.170486	Info	c2alice : c2alicesrv401	stagerd	24494	24504	File selected for deletion	ca827d26-cd71-4c34-96f9-4dfed9e9d259	-	**Filename**=/srv/castor/03/73/1207778773@castorns.14999559347 **NSHOSTNAME**=castorns **DiskServer**=lxfsrc1107.cern.ch
2013-01-08 16:17:08.298947	Info	c2alice : c2alicesrv201	stagerd	5017	5040	Request processed	4cf588ee-ec81-4492-bb72-0f09a40e2d9f	-	**Username**=root **SvcClass**=t0alice **NSHOSTNAME**=castorns **Filename**=/castor/cern.ch/alice/raw/global/2012/10/28/02/12000190903011.17.root **ProcessingTime**=0.020487 **Groupname**=root **SUBREQID**=d0bd8f9c-cea4-3cec-e043-46a18a895b60 **Type**=StagePrepareToGetRequest
2013-01-08 16:17:08.295364	Info	c2alice : c2alicesrv201	stagerd	5017	5040	Archiving subrequest	4cf588ee-ec81-4492-bb72-0f09a40e2d9f	-	**Username**=root **SvcClass**=t0alice **NSHOSTNAME**=castorns **Filename**=/castor/cern.ch/alice/raw/global/2012/10/28/02/12000190903011.17.root **Groupname**=root **SUBREQID**=d0bd8f9c-cea4-3cec-e043-46a18a895b60 **Type**=StagePrepareToGetRequest
									ClassId=0 **OwnerGid**=0 **Gid**=0 **Cwd**= **Function**=openx **ProcessingTime**=0.010 **ClientHost**=c2alicesrv201.cern.ch

c2adm01/logviewer/req_id/4cf588ee-ec81-4492-bb72-0f09a40e2d9f

Client monitoring

CERN IT Department

- Monitor also user action
  - Client version
    - Used for migration (security)
  - User activities
    - e.g. top users vs service class (abuses)
  - Protocols
    - Usage of RFIO, xroot, root, gFTP, SRM, … or mutable/immutable

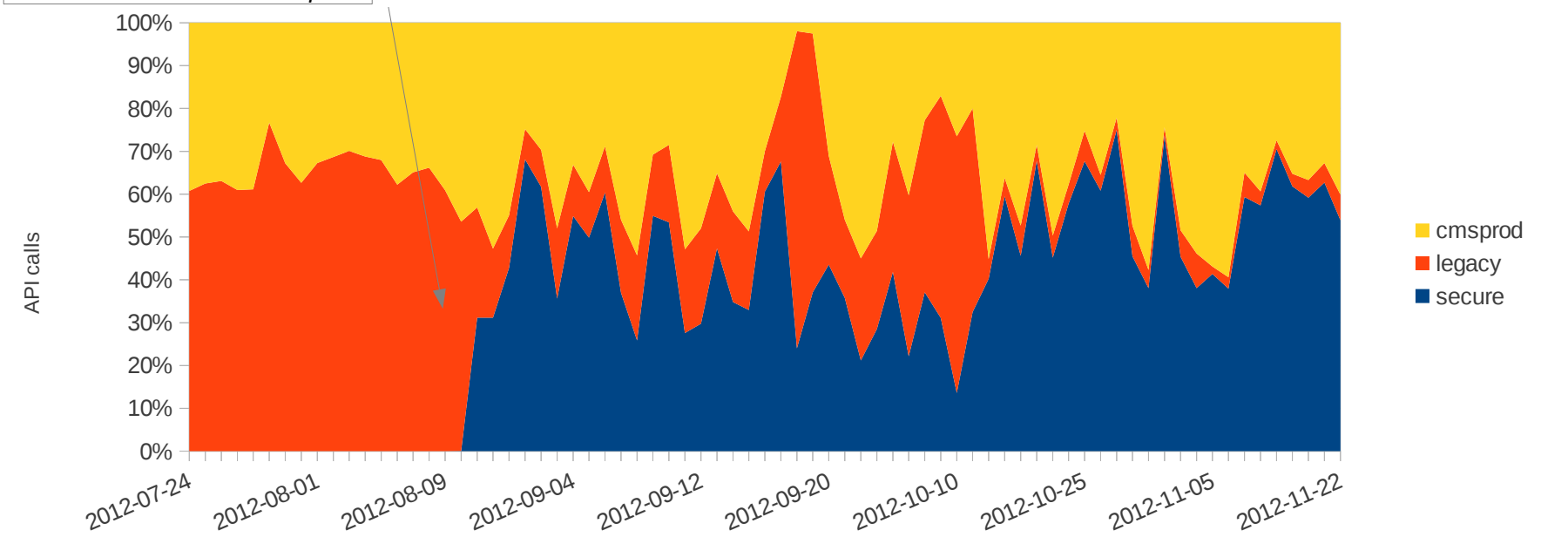


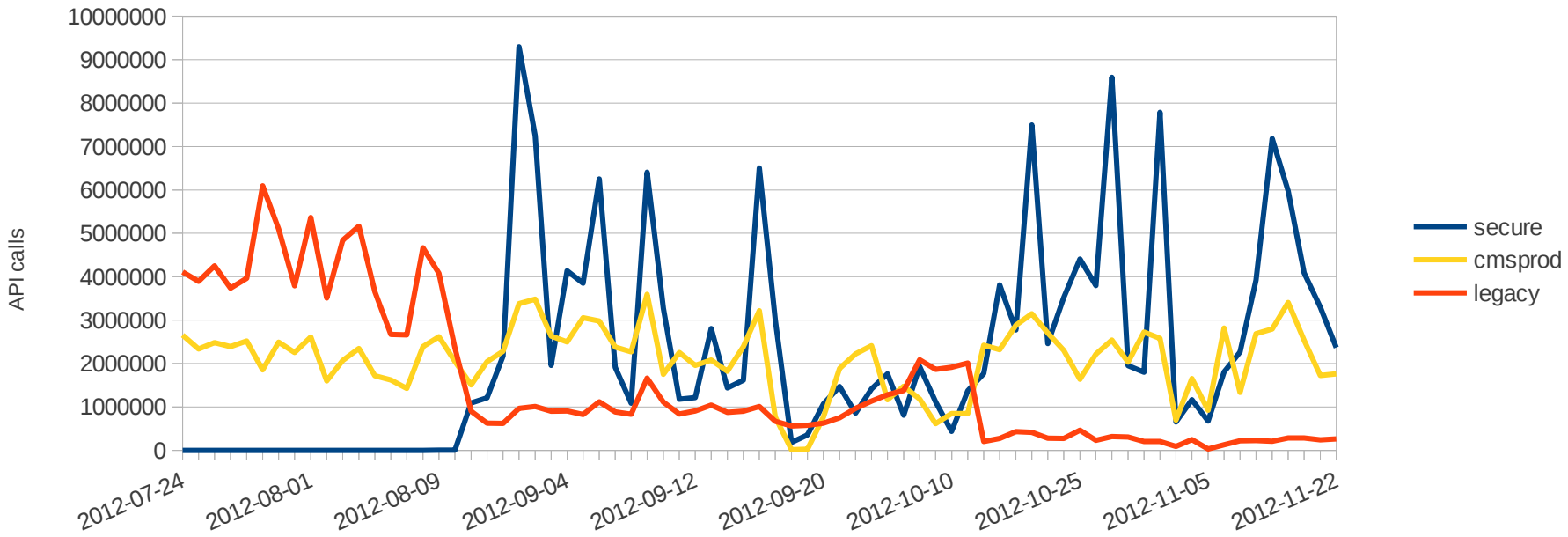
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Client-rollout – others than lx* and SLC nodes

CERN IT Department

ATLAS online code recompiled





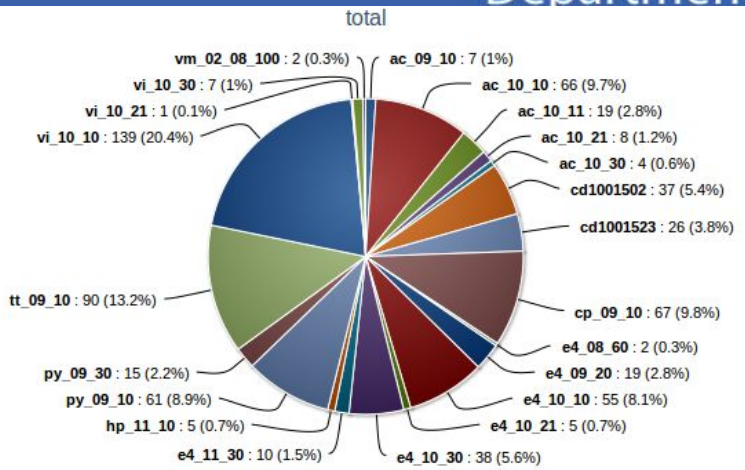
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Installation / HW break down / Machine status

castor

Filter :

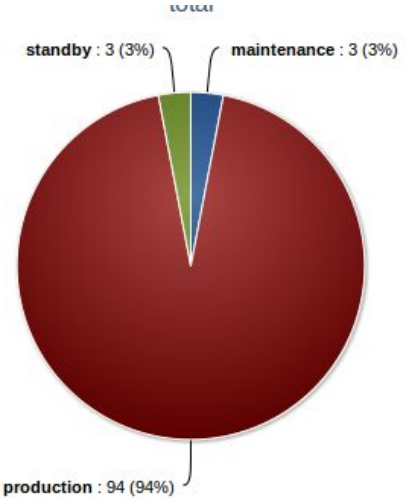
hwmodel	total	c2alice	c2atlas	c2cernt3	c2cms	c2lhcb	c2pps	c2public	c2repack
ac_09_10	7	1	1	2	1	1	0	1	0
ac_10_10	66	45	6	0	4	0	0	11	0
ac_10_11	19	0	0	0	12	0	0	7	0
ac_10_21	8	0	2	0	2	2	0	2	0
ac_10_30	4	0	1	0	0	0	0	3	0
cd1001502	37	8	14	0	11	4	0	0	0
cd1001523	26	6	6	2	12	0	0	0	0
cp_09_10	67	1	60	0	0	6	0	0	0
e4_08_60	2	0	0	0	0	0	2	0	0
e4_09_20	19	2	8	0	0	0	0	9	0
e4_10_10	55	0	3	0	6	4	1	41	0
e4_10_21	5	1	1	0	1	1	0	1	0
e4_10_30	38	33	0	0	0	0	0	5	0
e4_11_30	10	0	10	0	0	0	0	0	0
hp_11_10	5	1	1	0	1	1	0	1	0
py_09_10	61	0	46	0	4	5	2	4	0
py_09_30	15	0	1	0	2	1	0	11	0
tt_09_10	90	74	8	0	6	0	0	2	0
vi_10_10	139	7	37	0	38	22	1	34	0
vi_10_21	1	0	0	0	0	0	0	0	1
vi_10_30	7	0	0	0	0	7	0	0	0
vm_02_08_100	2	0	0	0	0	0	2	0	0
	Show	Show	Show	Show	Show	Show	Show	Show	Show
Total	683	179	205	4	100	54	8	132	1



c2cms

Filter :

status	total	archive	default	server	srm	t0export	t0input	t0streamer	t1transfer
maintenance	3	0	1	1	0	0	0	1	0
production	94	15	16	2	2	21	4	8	26
standby	3	0	0	0	0	0	0	3	0
	Show	Show	Show	Show	Show	Show	Show	Show	Show
Total	100	15	17	3	2	21	4	12	26



Other examples

- Nodes out of production for too long
  - SMS status (CERN tool to declare a node in production/standby/maintenance

for node in CASTOR:
    if not node.status=='production' and now-node.lastchange()>3days:
        → publish the name of the “bad guy” node

- “OutOfStager”
  - Nodes not giving heartbeat to the stager (headnodes) but being in production otherwise

for node in CASTOR:
    if node.status=='production' and node.stagerstatus=='DRAINING':
        → publish the name of the ”bad guy” node

```
msg = { 'EPOCH' : epoch, 'USECS' : usecs,'INSTANCE' : this_instance,
        'SvcClass' : this_svcclass,'TIMESTAMP' : c2timestamp,
        'outofstager_host' : host }
data_.append(msg)
```

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Example of another consumer (AI infrastructure)

CERN IT Department

GNI Dashboard | Browse notifications | sponce | Info

Refresh

From **now** to **7 days ago**

State : Open

Producer : All

Snow : All

Cluster :

≠

ms_win

Add an entity filter

Add a metric filter

Add a generic filter

Show 20 entries

	Created	Cluster	Entities	Metrics	Snow	Links
⊕	2013-04-12 10:20:04	nocontactprocessor	aimon03	no_contact	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lemonpupvm98	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lemonpupvm95	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	pedrotestlf	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lemonpupvm94	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lfcpt01	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lemonpupvm91	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lemonpupvm97	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	lemonpupvm96	no_contact_lemon	-	-
⊕	2013-04-11 11:52:32	ai-monitoring	c2adm01	no_contact_lemon	-	-
⊕	2013-04-09 14:14:35	nocontactprocessor	lemonpupvm89	no_contact	-	-
⊕	2013-04-09 14:11:35	nocontactprocessor	lemonpupvm92	no_contact	-	-
⊕	2013-04-09 14:08:35	nocontactprocessor	lemonpupvm90	no_contact	-	-
⊕	2013-04-05 18:52:30	aimon/spare	aimon02	exception.puppetd_wrong	-	▾
⊕	2013-04-05 17:35:16	nocontactprocessor	lemonpupvm91	no_contact	-	-
⊕	2013-04-05 17:35:16	nocontactprocessor	lemonpupvm97	no_contact	-	-
	Created	**Cluster**	**Entities**	**Metrics**	**Snow**	**Links**

Showing 16 entries (filtered from 5,087 total)

← Previous | 1 | Next →

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

HADOOP

- Able to put 10k log msg/s in HDFS
  - (msg size between 500 and 1000 bytes)
  - ~10MB/s
  - For CASTOR and EOS
    - 30-60 GB/day for each system
- Uses CERN Agile Infrastructure messaging

Cluster Summary

Security is *OFF*
6848700 files and directories, 1809174 blocks = 8657874 total.
Heap Memory used 1.39 GB is 84% of Commited Heap Memory 1.64 GB. Max Heap Memory is 5.21 GB.
Non Heap Memory used 45.29 MB is 99% of Commited Non Heap Memory 45.69 MB. Max Non Heap Memory is 130 MB.

Configured Capacity	:	624.4 TB			
DFS Used	:	237.19 TB			
Non DFS Used	:	32.21 TB			
DFS Remaining	:	355 TB			
DFS Used%	:	37.99 %			
DFS Remaining%	:	56.85 %			
Block Pool Used	:	237.19 TB			
Block Pool Used%	:	37.99 %			
DataNodes usages	:	Min %	Median %	Max %	stdev %
		23.19 %	32.71 %	51.54 %	11.19 %
Live Nodes	:	16 (Decommissioned: 1)			
Dead Nodes	:	5 (Decommissioned: 2)			
Decommissioning Nodes	:	0			
Number of Under-Replicated Blocks	:	0			

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

HADOOP

- Data to HADOOP (HDFS)
  - Log repository of CASTOR/EOS logs independent from CASTOR/EOS (disaster recovery)
    - Recover (reconstruct) file metadata in case of disaster recovery
  - Several TBs (~300+ days of history)
- Analysis on HADOOP (MapReduce)
  - Interesting technology
  - (Almost) in production
    - Systematic studies using “new” metrics on “old” data (the cockpit works 'on the flight'
    - (Re)compute data to be fed in the cockpit

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Where are we?

- In production with CASTOR
- To enter production for EOS soon
- As more (non CASTOR/EOS) information enter the system it will become more powerful
  - Analysis/Presentation/Correlation
    - Inter-service correlations (batch farm CPU usage vs disk server farm IO)
  - “Generic” components entering full production soon (e.g. Alarms see previous slide)
- Improve the feedback loop Hadoop to Cockpit
  - simplify the historical searches using Cockpit metrics ran by MapReduce
- Extend Hbase usage?
  - e.g. simplify access to filtered informations (KPI)

Effort/contacts

- Technical leadership of the DSS Cockpit (and coordination with the CERN AI monitoring project): S. Ponce
- Very effective technical students in our team: (Stefano Russo), (Benjamin Fiorini), Manuel Servais
  - Part of their thesis work: () → contract finished and thesis done
- Nice collaboration with the CERN AI monitoring project (P. Andrade et al.), AI services in operations (Apollo, Hadoop,...) operated by different IT groups
- CASTOR/EOS operation team and all the FDO section for suggestions/feedback/contributions

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Service manager view

- **Lessons learnt (or simply refreshed)**
  - The data are “somewhere” does not mean you are using them as you should!
  - Correlation (not only time series) is essential
  - KPIs are the “champions” of your (larger and ever improving) set of day-by-day plots/table/etc... “indicators”
  - **Sometimes enabling users to peek into the system improves our service more than coding new functionality (understanding, trust relation)**
- **Outlook**
  - We've got the data: no excuse we should use them! ;)
    - Analysis (i.e. gets our hands dirty analysing them)
  - **Do we quantitatively understand monitoring data correlations?**
    - Some hints in the presentation of Luca Mascetti
    - E.g.to feed back in the next market survey or hardware acceptance?
  - Can we have an early warning system?
    - Service will be down in 100 hours?
      - Unless user xyz stops bombarding disk pool abc?
      - (Uncharted) machine-learning world

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Questions?

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it