An Overview of Operating Systems, Installation Procedures and Monitoring at CSCS

Peter Öttl

on behalf of Pablo Fernandez, Fotis Georgatos, Jason Temple and Peter Öttl, HCS Team

Outline

- Operating Systems for the Swiss WLCG Tier-2

- Installation Methods Chosen

- Configuration Management

- Monitoring Tools

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CSCS
Swiss National Supercomputing Centre

The Phoenix LCG Tier-2 Cluster, PhaseB

- **SLC4 (32-bit)**
  - 1 CE
  - 1 MON-Box
  - 1 User Interface
  - 3 site-BDII's

- **SL5 64-bit**
  - 60 WN's

- **CentOS 5.x**
  - 2 Storage Head Nodes
  - 2 NFS servers
  - 3 XEN Dom0's

- **Solaris 10**
  - 20 Storage Servers

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The Phoenix LCG Tier-2 Cluster, PhaseB

- Ideally only one OS, but: requirements from gLite middleware are SL(C)4 and SL5

- SL4
  - SL4.0 - 4.8: Obsolete in October 2010
  - SL4.9: Minimal support (security only) until February 2012
    - Legacy release (trimmed down)
    - No release date yet

- SL5
  - SL5.4: Support until February 2012
  - SL5.5: No release date yet

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The OS dilemma

- Solaris/OpenSolaris vs. Linux

  - Pro:
    - ZFS
    - we already use it

  - Contra:
    - Performance
    - Updates have proven to be pain
    - Support (and costs, even for OpenSolaris updates)

  - Opt-out
    - Incompatible with our IB-Ofed stack + Mellanox bridge

The Phoenix LCG Tier-2 Cluster, PhaseC

- ## SL5.4
  - For Grid services only
  - Tradeoff: lcg-CE supported by gLite only in SL4
    - SL5 possible, but needs extra work; we picked this choice

- ## RHE5.4
  - Everything else:
    - Lustre
    - Xen Dom0's
    - Etc.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Installation Methods

- **Physical Machines**
  - PXE boot with kickstart
  - cfengine is being setup by a script
  - cfengine does the rest (of the magic)

- **Virtual Machines**
  - Copy a master image to disk+mkswap (both lvm volumes)
  - Script configures xen image with network parameters
  - cfengine is being setup by a script
  - cfengine does the rest (of the magic)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

2nd HPC-CH Workshop
19th May 2010, EPF Lausanne

CSCS
Swiss National Supercomputing Centre

7

Configuration Management

- ## OS: we inherited cfengine
  - works for us
  - but stuck with version 2 and will evaluate alternatives

- ## gLite Middleware
  - done by YAIM (Yet Another Installation Manager)
  - assisted by cfengine to complete the work

- ## Updates of software
  - done manually, on purpose
    - may break middleware; need for a preproduction environment

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

RedHat Satellite Network

- License available through ETHZ
- Direct support from RedHat allows for timely security updates vs. other OS's, such as CentOS or SL
- Simple to automate and install, even with kickstart installations

Client Registration

Folgendes Skript dient dazu ein bereits installiertes System am RHN zu registrieren. Dabei spielt es keine Rolle ob es sich um ein neu installiertes oder von der JEDEM FALL das gesamte Script aus. $KEY sollte bei der Verwendung durch einen gültigen Activation Key ersetzt werden.

Hinweis: Beachten Sie bitte das dem Activation Key eine Systemgruppe zugeordnet sein muss auf die Sie Zugriff haben, um das System über den Satellite Ser

```
# Diese Dienste muessen fuer rhnreg_ks laufen, wenn rhn-setup aelter als 0.4.20-33.el5_5.1
if [[ $(rpm -qa rhn-setup --queryformat '%{VERSION}-%{RELEASE}') < "0.4.20-33.el5_5.1" ]]; then

  service messagebus status
  if [ $? -ne 0 ]; then
    service messagebus start
  fi

  service haldaemon status
  if [ $? -ne 0 ]; then
    service haldaemon start
  fi

fi

# evtl. Deinstallation alter SSL Zertifikate
rpm -e rhn0-ethz-trusted-ssl-cert > /dev/null 2>&1
rpm -e rhn-register-ethz > /dev/null 2>&1

# Installation der neuen Zertifikate
rpm -Uhv http://id-rhns-prd.ethz.ch/pub/rhn-org-trusted-ssl-cert-1.0-2.noarch.rpm

# Loeschen der alten osad Authentifizierungsinformationen (fuer re-registrierung notwendig)
rm -f /etc/sysconfig/rhn/osad-auth.conf > /dev/null 2>&1

# Migration von alter zu neuer Infrastruktur: alte Proxies/alter Satellite -> neuer Satellite
perl -npe 's/rhn0.ethz.ch/id-rhns-prd.ethz.ch/g' -i /etc/sysconfig/rhn/up2date
perl -npe 's/rhn1.ethz.ch/id-rhns-prd.ethz.ch/g' -i /etc/sysconfig/rhn/up2date
perl -npe 's/redhat.ethz.ch/id-rhns-prd.ethz.ch/g' -i /etc/sysconfig/rhn/up2date

# Neuinstallation: RHN Hosted Zertifikat -> RHN Satellite Zertifikat und RHN Hosted -> neuer Satellite
perl -npe 's/RHNS-CA-CERT/RHN-ORG-TRUSTED-SSL-CERT/g' -i /etc/sysconfig/rhn/up2date
perl -npe 's/xmlrpc.rhn.redhat.com/id-rhns-prd.ethz.ch/g' -i /etc/sysconfig/rhn/up2date
```

RedHat Satellite Network

RedHat Satellite Network - Management

- Allows for ease of management - the package update notification can be sent via e-mail, and it is easy to schedule errata updates for your systems

RED HAT NETWORK SATELLITE

| Systems | | Search |

Overview | **Systems** | **Errata** | **Channels** | **Schedule** | **Help**

5 SYSTEMS SELECTED | MANAGE | CLEAR

Overview
Systems
System Groups
System Set Manager
   Status
Advanced Search
Activation Keys
Stored Profiles

Errata Legend

🔒 Security
🐞 Bug Fix
⚡ Enhancement

System Set Manager ⁇

Overview Systems Errata Packages Channels Configuration Misc

Relevant Errata List

The errata below apply to one or more of your selected systems.

1 - 25 of 235 (218 selected) |◁ ◁ ▷ ▷|

	Type	Advisory	Synopsis	Systems	Updated
☑	🐞	RHBA-2010:0415	device-mapper-multipath bug fix update	5	2010-05-17
☐	🔒	RHSA-2010:0398	Important: kernel security and bug fix update	5	2010-05-06
☑	🐞	RHBA-2010:0390	mkinitrd bug fix update	5	2010-05-04
☑	⚡	RHEA-2010:0381	tzdata enhancement update	5	2010-04-28
☑	🔒	RHSA-2010:0382	Important: xorg-x11-server security update	5	2010-04-28
☑	🐞	RHBA-2010:0375	kexec-tools bug fix update	5	2010-04-23
☑	🐞	RHBA-2010:0368	lvm2 bug fix update	5	2010-04-22
☑	🔒	RHSA-2010:0361	Moderate: sudo security update	5	2010-04-20
☑	🔒	RHSA-2010:0347	Moderate: nss_db security update	5	2010-04-13
☑	🐞	RHBA-2009:1605	tcpdump bug fix update	5	2010-04-07
☑	🐞	RHBA-2009:1629	OpenIPMI bug fix update	5	2010-04-07
☑	🔒	RHSA-2010:0343	Important: krb5 security and bug fix update	5	2010-04-06
☑	⚡	RHEA-2010:0336	tzdata enhancement update	5	2010-03-31
☑	🐞	RHBA-2010:0294	xen bug fix and enhancement update	5	2010-03-30
☑	🐞	RHBA-2010:0179	kexec-tools bug fix update	5	2010-03-30

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CSCS
Swiss National Supercomputing Centre

RED HAT NETWORK SATELLITE

Systems ▼ [   ] Search

| Overview | Systems | Errata | Channels | Schedule | Help |

5 SYSTEMS SELECTED MANAGE CLEAR

Overview
Systems
System Groups
System Set Manager
  Status
Advanced Search
Activation Keys
Stored Profiles

Errata Legend
🔒 Security
🐞 Bug Fix
⚡ Enhancement

System Set Manager ❓

Overview   Systems   Errata   Packages   Channels   Configuration   Misc

Relevant Errata List

The errata below apply to one or more of your selected systems.

1 - 25 of 235 (218 selected) |◁ ◁ ▷ ▷|

☐	Type	Advisory	Synopsis	Systems	Updated
☑	🐞	RHBA-2010:0415	device-mapper-multipath bug fix update	5	2010-05-17
☐	🔒	RHSA-2010:0398	Important: kernel security and bug fix update	5	2010-05-06
☑	🐞	RHBA-2010:0390	mkinitrd bug fix update	5	2010-05-04
☑	⚡	RHEA-2010:0381	tzdata enhancement update	5	2010-04-28
☑	🔒	RHSA-2010:0382	Important: xorg-x11-server security update	5	2010-04-28
☑	🐞	RHBA-2010:0375	kexec-tools bug fix update	5	2010-04-23
☑	🐞	RHBA-2010:0368	lvm2 bug fix update	5	2010-04-22
☑	🔒	RHSA-2010:0361	Moderate: sudo security update	5	2010-04-20
☑	🔒	RHSA-2010:0347	Moderate: nss_db security update	5	2010-04-13
☑	🐞	RHBA-2009:1605	tcpdump bug fix update	5	2010-04-07
☑	🐞	RHBA-2009:1629	OpenIPMI bug fix update	5	2010-04-07
☑	🔒	RHSA-2010:0343	Important: krb5 security and bug fix update	5	2010-04-06
☑	⚡	RHEA-2010:0336	tzdata enhancement update	5	2010-03-31
☑	🐞	RHBA-2010:0294	xen bug fix and enhancement update	5	2010-03-30
☑	🐞	RHBA-2010:0179	kexec-tools bug fix update	5	2010-03-30

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

2nd HPC-CH Workshop
19th May 2010, EPF Lausanne

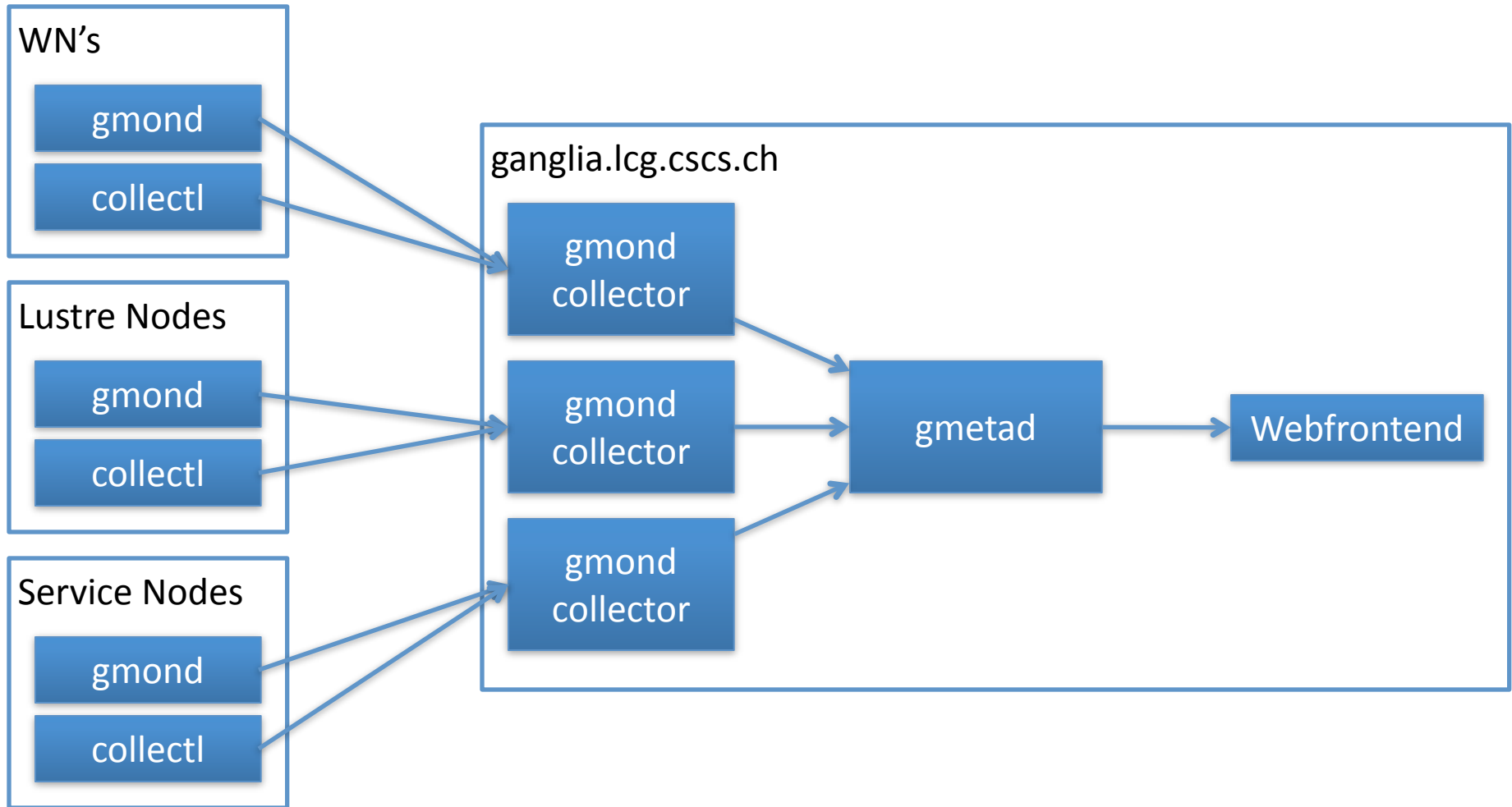CSCS
Swiss National Supercomputing Centre

13

Monitoring of IB traffic & Lustre read/writes

- Ganglia is not monitoring IB traffic and Lustre I/O

- Even ifconfig does not see the "raw" IB interface
  - no simple commands to monitor the flow of IB packets

- Collectl provides lightweight collection of device performance information including IB and Lustre
  - support to export to Ganglia > 3.1.x

Collectl Metrics

- Infiniband
  - KBIn, PktIn
  - KBOut, PktOut

- Lustre
  - Lustre Client
    - Errs
    - KBRead, Reads
    - KBWrite, Writes
  - Lustre FS single OST statistics and OST summary
    - KBRead, Reads
    - KBWrite, Writes
  - Lustre MDS
    - Gettatr, GttrLck, StatFS, Sync, Connect, Create, Settatr, Rename,etc.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CSCS
Swiss National Supercomputing Centre

Collectl integration to Ganglia

Setup collectl & Ganglia

- **Easy part: Setup collectl**
  - /etc/collectl.conf
    - on OSS nodes:
      DaemonCommands = -f /var/log/collectl -slLdx --export gexpr,ganglia:8691,i=20 -r00:01,7 -m -F60
    - on MDS nodes:
      DaemonCommands = -f /var/log/collectl -sldDx --export gexpr,ganglia:8691,i=20 -r00:01,7 -m -F60
    - on WNs nodes:
      DaemonCommands = -f /var/log/collectl -slx --export gexpr,ganglia:8690,i=20 -r00:01,7 -m -F60

- **Hard part: Setup Ganglia**
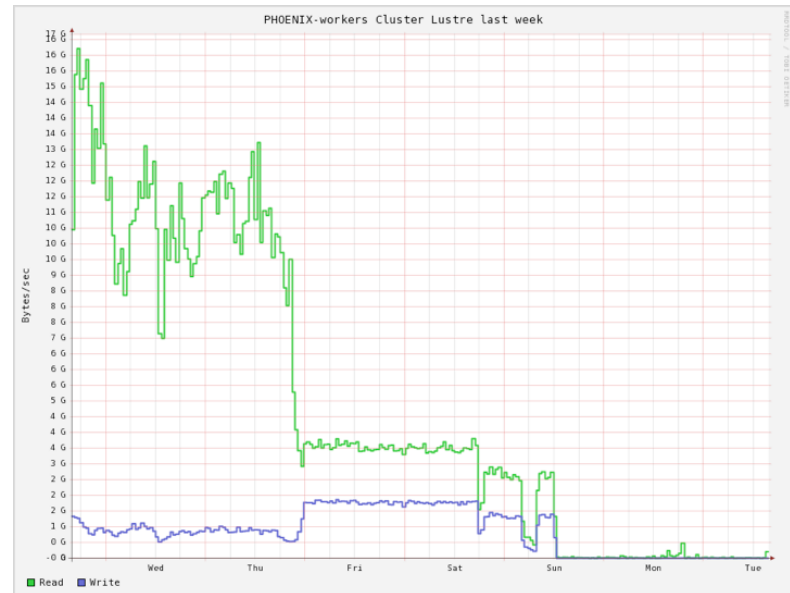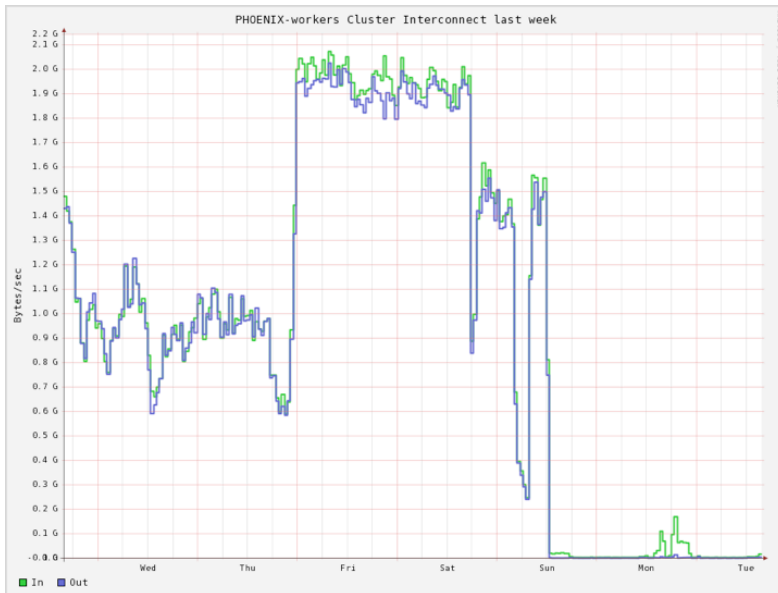 - no RPMs available for recent versions
    - Need to compile from source
  - lots of dependencies
  - lots of configuration files

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Ganglia Webfrontend Modifications

- To get aggregate over a cluster need to add "graph handlers"
  - http://sourceforge.net/apps/trac/ganglia/wiki/ganglia_modular_graph_documentation

References

- Ganglia
  - http://ganglia.sourceforge.net

- collectl
  - http://collectl.sourceforge.net