Quattor GRIF Example

Michel Jouvin
jouvin@lal.in2p3.fr
Quattor Tutorial
LCG T2 Workshop
CERN, June 16 2006

GRIF at a Glance

- GRIF : a collaborative effort between 5 HEP labs in Paris Region to build a LCG T2 for analysis and MC
  - 4 LHC experiments (80%) + other EGEE/local VOs (20%
  - Planned resources in 2008 : 3000 CPUs, 700 TB
  - A unique grid site geographically distributed (5 locations

- Quattor choosen in 2005 to be a unique deploymen and configuration tool for the whole GRIF
  - Based on LAL experience with Quattor
  - A unique configuration database to warrant consistency and leverage administration
  - Other local usage possible at each site

GRIF Numbers

- 5 grid « sites » sharing the same base configuration
  - DAPNIA : 20
  - LAL : 40
  - LPNHE : 15
  - LLR : 5
  - IPNO : 10
  - 1 non grid site at CEA : Orme
  - All services managed by Quattor (SPMA) : CEs, SEs, RB, VOMS, MON, MYPROXY...
- Non grid usage at LAL
  - Non grid servers for physics, web servers... : 35
    - All servers for physics (15) configured as LCG UI
    - Whole configuration managed by Quattor/SPMA
  - Desktops and virtual machines : 5 (more to come)
    - SPMA configured to allow for user installed RPMs
  - Share basic configuration with LAL grid nodes

Quattor Configuration

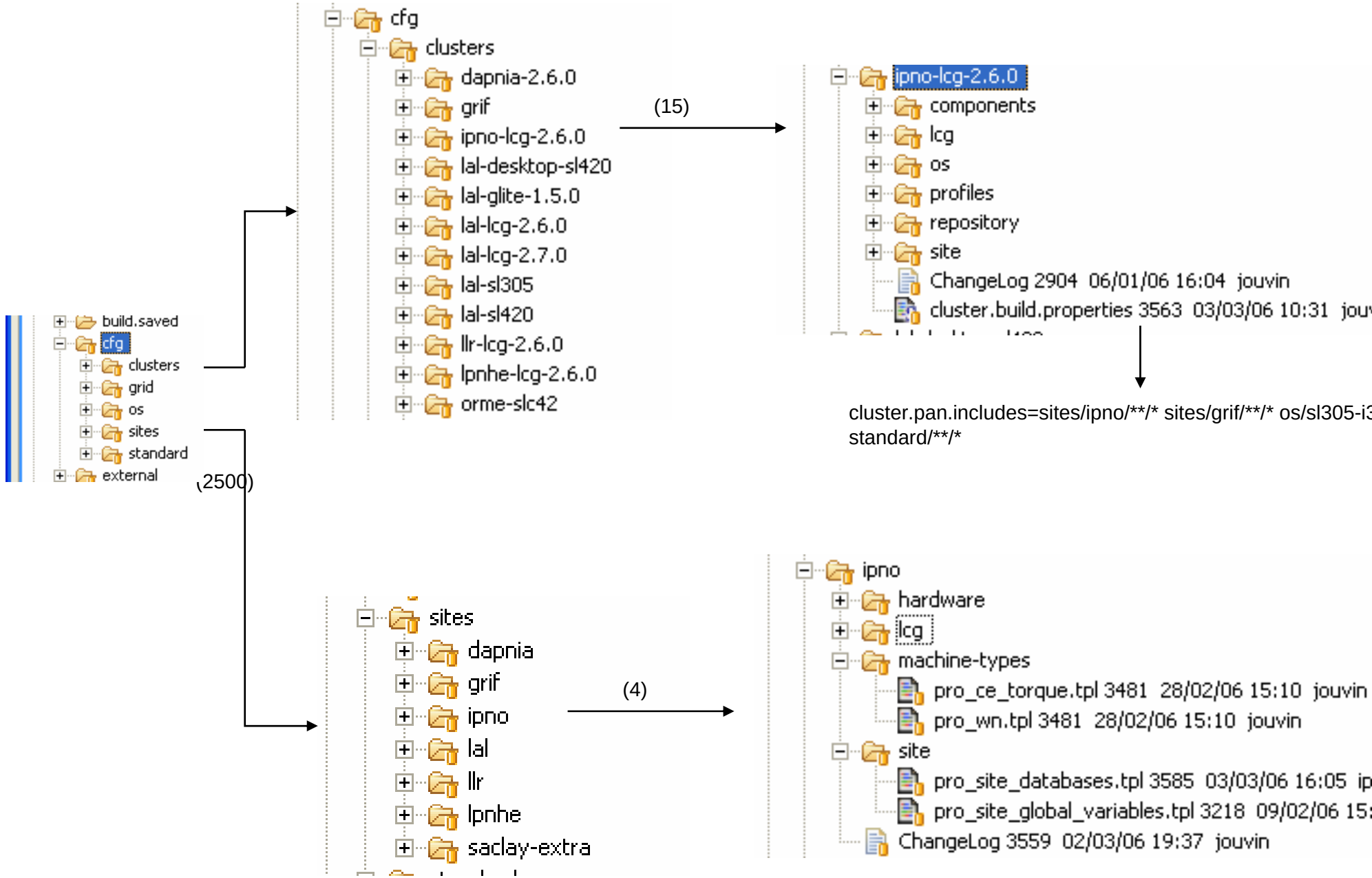
- Configuration Database based on Subversion (SCDB
  - Hosted by LAL Subversion server
    - Distinct from Quattor server
- SW repository : based on httpd
- Initial installation with AII for all Quattor managed nodes
  - LAL : DHCP/Tftp servers on Quattor server are shared w non Quattor usage (main servers for LAL)

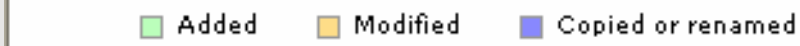
In Quattor Configuration

- 1 Quattor server master at LAL
  - AII related services (DHCP + TFTP)
  - Repository server (http)
  - Machine profile server : recompile templates from SCDB at each configuration deployment (ant deploy)
    - Deployment can occur only after successful local compilation
    - Incremental compilation
- 1 Quattor server per geographical site
  - Currently only used for AII (DHCP+TFTP)
  - Plan to serve template for each site : parallel compilation at each site ?
  - Repository replicas ? SWrep or http cache?
- Full recompile = 3 mn (#150 nodes)
  - Quattor server = dual CPU Opteron 2,2 Ghz
  - 5 mn considered as a maximum (plan to x5 machines)

Template Layout

- Based on standard layout
  - See presentation on Quattor Fabric Description
- Specificities related to our multi-site configuration
  - Minimize the number of site specific templates
  - Allow each site to specify its site specific templates (e.g. network parameters) without duplication
- ‘sites’ hierarchy added
  - set of common parameters shared at a geographical location or a service entity
  - A cluster can belong to one or several sites
    - e.g. : LAL, GRIF
  - GRIF site is non geographical : all parameters common t all GRIF sites (repositories...)


build.saved
cfg
clusters
grid
os
sites
standard
external
(2500)
cfg
clusters
dapnia-2.6.0
grif
ipno-lcg-2.6.0
lal-desktop-sl420
lal-glite-1.5.0
lal-lcg-2.6.0
lal-lcg-2.7.0
lal-sl305
lal-sl420
llr-lcg-2.6.0
lpnhe-lcg-2.6.0
orme-slc42
(15)
ipno-lcg-2.6.0
components
lcg
os
profiles
repository
site
ChangeLog 2904 06/01/06 16:04 jouvin
cluster.build.properties 3563 03/03/06 10:31 jouv
cluster.pan.includes=sites/ipno/**/* sites/grif/**/* os/sl305-i3
standard/**/*
sites
dapnia
grif
ipno
lal
llr
lpnhe
saclay-extra
(4)
ipno
hardware
lcg
machine-types
pro_ce_torque.tpl 3481 28/02/06 15:10 jouvin
pro_wn.tpl 3481 28/02/06 15:10 jouvin
site
pro_site_databases.tpl 3585 03/03/06 16:05 ip
pro_site_global_variables.tpl 3218 09/02/06 15:
ChangeLog 3559 02/03/06 19:37 jouvin



View changes

Old	New		Date	Rev	Chgset	Author	Log Message
○	◉		03/06/06 14:12:07	@3599	[3599]	jouvin	Fix disk information for grid27
◉	○		03/06/06 14:02:11	@3597	[3597]	jouvin	Fix disk information for grid27
○	○		03/03/06 18:42:31	@3596	[3596]	lpnhe	Modifications following the last updates in the Grif site structure at ...
○	○		03/03/06 17:51:31	@3594	[3594]	ipno	modif pro_hardware_machine..hp_proliant..
○	○		03/03/06 17:41:04	@3591	[3591]	ipno	modif pro_hardware_machine..hp_proliant..
○	○		03/03/06 17:35:47	@3588	[3588]	ipno	modif pro_hardware_machine..hp_proliant..
○	○		03/03/06 16:05:39	@3585	[3585]	ipno	ajout ipnsedpm.in2p3.fr
○	○		03/03/06 15:35:16	@3583	[3583]	jouvin	Move grid27 to SL4.2 test cluster
○	○		03/03/06 15:07:19	@3580	[3580]	jouvin	Rename cluster orme-slc42 slc/ to machine-types/, remove previous ...
○	○		03/03/06 14:57:54	@3579	[3579]	jouvin	Rename cluster orme-slc42 lal/ to machine-types/
○	○		03/03/06 14:37:02	@3577	[3577]	jouvin	Definition of machine type pro_lal_desktop
○	○		03/03/06 14:36:01	@3576	[3576]	jouvin	Definition of machine type pro_lal_desktop

Changeset 3568

Timestamp: 03/03/06 11:14:27
Author: jouvin

Message: Use an OS version independent template to add openafs client; define auger1 as an xtremweb server

Files:
- trunk/cfg/clusters/lal-sl420/profiles/profile_auger1.tpl (1 diff)
- trunk/cfg/os/sl305-i386/os/pro_os_openafs_client.tpl
- trunk/cfg/os/sl420-i386/os/pro_os_openafs_client.tpl
- trunk/cfg/sites/lal/machine-types/pro_lal_config_afs_client.tpl (1 diff)

Unmodified | Added | Removed | Modified | Copied | Moved

trunk/cfg/clusters/lal-sl420/profiles/profile_auger1.tpl

r3516	r3568	
10	10	
11	11	`define variable XW_STARTUP_START = false;`
12		`#include pro_lal_server_physics_xtremweb;`
13		`include pro_lal_server;`
	12	`include pro_lal_server_physics_xtremweb;`
14	13	
15	14	

trunk/cfg/sites/lal/machine-types/pro_lal_config_afs_client.tpl

r3371	r3568	
12	12	`define variable PKG_ARCH_KERNEL_MODULE_OPENAFS = PKG_ARCH_KERNEL;`
13	13	

Terminé

Conclusions

- Very happy with Quattor...!!!
  - SCDB allows to set up a very flexible system : ACLs, We interfaces, notification (RSS)...
    - Site adminstrators restricted to modify their site/cluster templates
  - Template layout allows simultaneous support for very different configurations
  - Management from various platform in disconnected mod
- Main concern is pan compiler speed
  - Not affordable to spend more than 5 minutes to recompile everything
  - 1 pre-compiled template for every worker node with just a few “last minute” customization
  - Parallel compilation of every site or on several CPUs
  - Rewrite of panc in Java to improve memory managemen
- Documentation effort (in french), based on Trac
  - http://trac.lal.in2p3.fr/Quattor