ELFms industrialisation plans

CERN openlab workshop

13 June 2005

German Cancio

CERN IT/FIO

http://cern.ch/elfms

Outline

- Background
- What is ELFms
- Collaboration with industry
- Challenges

Background

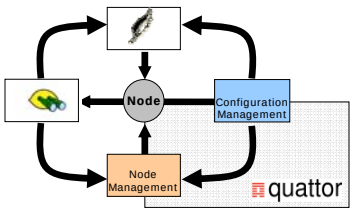
- A collaboration agreement is being worked out between CERN and an European SME
  - Leading provider of cluster systems
- Aim: further development of CERN's Extremely Large Fabric management system (ELFms) in an industrial context
- Example of how a software toolsuite initiated by EGEE's predecessor (EDG) is developed, deployed … and now being industrialised
  - Inspiration for potential EGEE spin-offs?

Fabric Management with ELFms

ELFms stands for '**E**xtremely **L**arge **F**abric **m**anagement **s**ystem'

Subsystems:

- quattor : configuration, installation and management of nodes
- : system / service monitoring
- : hardware / state management



- ELFms manages and controls most of the nodes in the CERN CC
  - ~2400 nodes out of ~ 3100
  - Multiple functionality and cluster size (batch nodes, disk servers, tape servers, DB, web, …)
  - Heterogeneous hardware (CPU, memory, HD size,..)
  - Supported OS: RH Linux / Scientific Linux (on i386,ia64,x86_64) and Solaris 9

quattor

http://quattor.org

Quattor

quattor

Quattor takes care of the *configuration, installation* and *management* of fabric cluster and nodes

➔A **Configuration Database** holds the 'desired state' of all fabric elements

- Node setup (CPU, HD, memory, software RPMs/PKGs, network, system services, location, audit info…)
- Cluster (name and type, batch system, load balancing info…) and site-wide attributes
- Defined in templates arranged in hierarchies – common properties set only once

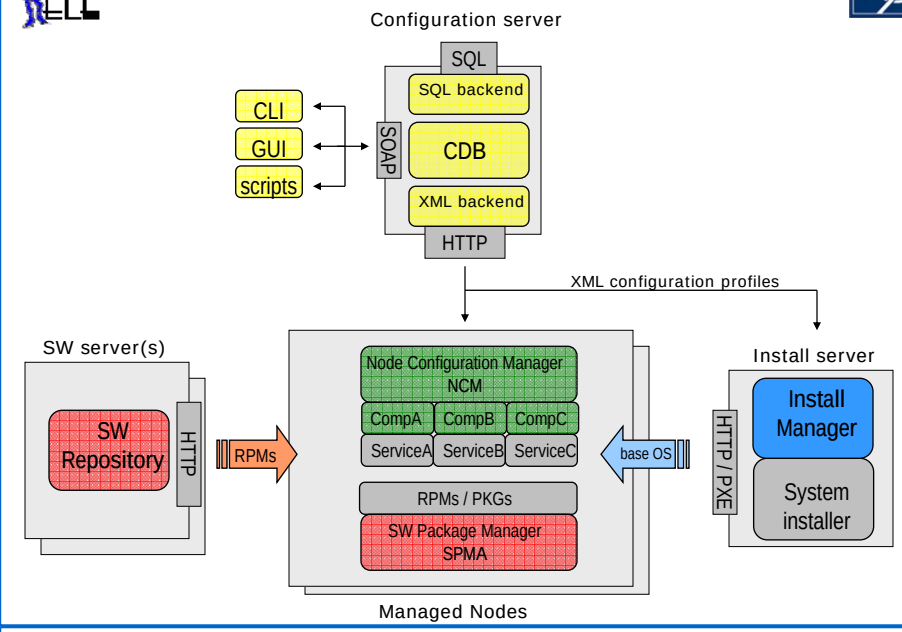
➔**Autonomous management agents** running on the node for

- **Base installation**
- **Service (re-)configuration**
- **Software installation and management**

- Quattor addresses heterogeneity
  - Functionality
  - Platforms/OS
  - Hardware
- Quattor addresses scalability
  - Management of O(10K) nodes with proxy infrastructure

Architecture

quattor



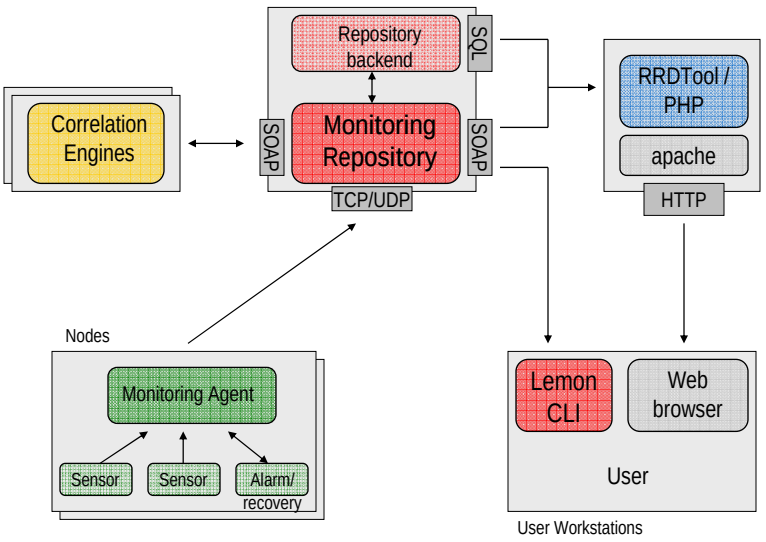
Lemon

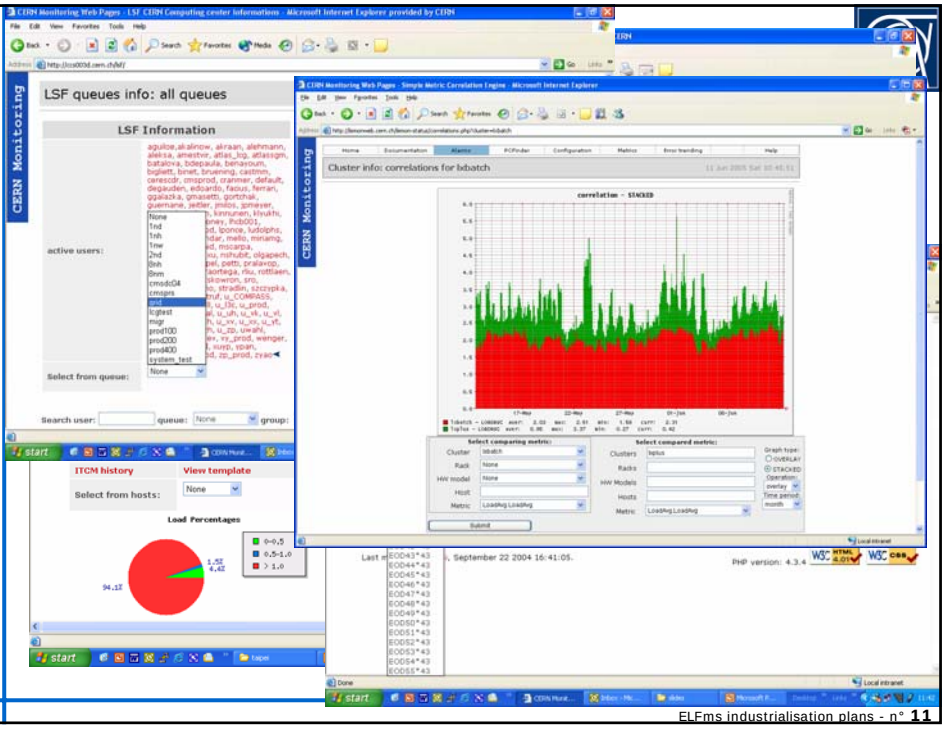
http://cern.ch/lemon

Lemon – LHC Era Monitoring

- *Performance* and *exception monitoring* of nodes and clusters
- Distributed system, scalable to O(10K) nodes
- Provides active monitoring of software and hardware
  - Facilitates early error detection and problem prevention
- Executes corrective actions
  - Global and local correlation and fault tolerance engines
- Provides persistent storage of monitoring data
  - Oracle or flat-file based back-end holding all event history

Architecture





leaf

http://cern.ch/leaf

LEAF - LHC Era Automated Fabric

- LEAF is a collection of workflows for *high level* node hardware and state management, on top of Quattor and LEMON:
- HMS (Hardware Management System):
  - Track systems through all *physical* steps in lifecycle eg. installation, moves, vendor calls, retirement
  - Automatically requests installs, retires etc. to technicians
  - GUI to locate equipment physically
- SMS (State Management System):
  - Automated handling (and tracking of) high-level configuration steps
    - E.g. progressively reconfigure / reboot cluster nodes for new kernel and/or physical move
- LEAF implementation is CERN specific, but concepts and design should be generic

Development / Deployment

- ELFms (Quattor/Lemon) were started in the scope of EU DataGrid.

- Development is now coordinated by CERN/IT in collaboration with other HEP institutes

- Quattor/Lemon are used in production in/outside CERN
  - LCG T1/T2 sites, ranging from 50-800 nodes/site
  - Complete configuration of system and LCG Grid middleware via Quattor
  - Integration with Grid services e.g. monitoring (GridICE, MonALISA)

Collaboration with Industry

- Collaboration between CERN and European SME partner

Mutual interest:

- SME in ELFms:
  - Profit from innovations in Quattor/Lemon architecture and design, in particular scalability and heterogeneity
  - Extend it for remote (WAN), secure, management of sites (University institutes, company offices, etc)
  - Make a product out of it: easy-to-use interfaces, setup wizards, advanced GUI's; integrate with own HPC technology and tools
- ELFms in SME:
  - Leverage knowledge in GUI's, easy-to-use interfaces, setup wizards
  - Remote management of sites (LCG T2 sites, online experiments)
  - Technology Transfer: demonstrate returns to industry of research results

Collaboration with Industry (II)

- Framework: **P**PARC's **I**ndustrial **P**rogramme **S**upport **S**cheme (PIPSS)
  - Grant scheme for collaboration between UK industry and researchers in HEP and related fields
- Duration: 2 years
- Manpower: 2 FTE / year
  - 1 FTE funded by PPARC, another FTE coming from SME
- Work Packages:
  - WP1. Technology transfer of existing ELFms technology
  - WP2: Development of secure remote management functionality
  - WP3: Capabilities for commercial product

Collaboration with Industry (III)

- Exploitable results:
  1. Freely-available (open source) version of Quattor/Lemon with advanced functionality
  2. Commercially supported version of Quattor/Lemon, integrated with “added-value” proprietary extensions by SME
  3. Use of Quattor/Lemon within a major european engineering company (SME customer), setting up an internal Grid encompassing clusters at multiple geographical sites
- Status
  - Research Proposal (contains actual project milestones and deliverables) finished
  - Collaboration agreement being worked out

Challenges

- Evolution of core ELFms software
  - How to ensure a compatible evolution of “core” software and commercial extensions
  - Definition of standards and stable API’s
- Software licensing schema, IP rights
  - Original EDG software: BSD-like license
  - Current ELFms continues with EDG license
  - Licensing schema for (part of) collaboration results may be different
- ELFms collaborating institutes
  - Define relationships and re-define responsibilities
  - Cultural differences – HEP vs. industry