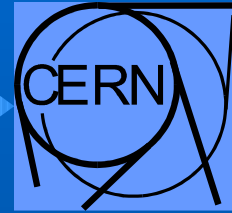


# HEPiX Large Cluster SIG Report

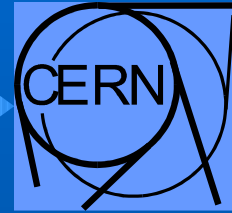
**Alan Silverman**  
**19<sup>th</sup> April 2002**  
**HEPiX 2002, Catania**

# Overview of the talk



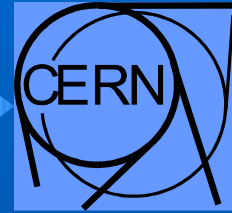
- **Cluster Builder Workshop**
- **Software Certification**
- **Large Site Operations**
- **OpenAFS**
- **Plans**

# Cluster Builders Workshop - 1



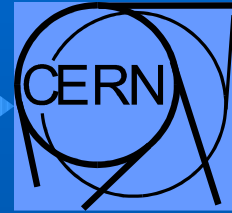
- A workshop to write *the* definitive guide to building and running a cluster - how to choose/select/test the hardware; software installation and upgrade tools; performance mgmt, logging, accounting, alarms, security, etc, etc
- Then document what exists and what might scale to large clusters.
- And by implication, what does not scale
- Maintain this.

# Cluster Builders Workshop - 2



- May 22<sup>nd</sup> to 25<sup>th</sup> 2001 in Fermilab
- Participation targetted at sites with minimum size (100-200 node) clusters
- Invitations sent, not only to HEP sites but other sciences, including biophysics. Also inviting participation by representatives from commercial firms with technical backgrounds
- 60 people attended

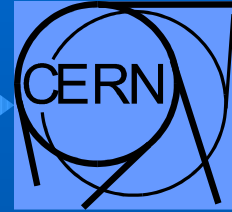
# Cluster Builders Workshop - 3



- Interesting presentations and discussions (workshop style)
- Proceedings published (including extra input)
- Summary presented at CHEP
- Started working on the actual Guide
- Interest expressed in a progress report style meeting late 2002/early 2003
- Proceedings and more information on the web

<http://conferences.fnal.gov/lccws/>

# Software Certification



- List of modules and products drawn up for Linux and for Solaris
- First entries submitted by FNAL and by CERN
- Published on HEPiX web site
- Due for an upgrade but delayed by re-organisation of the CERN IT Division!
- See link  
<http://wwwinfo.cern.ch/hepix/cluster/Linux-Cert-table>  
for the Linux example

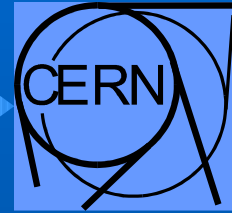
# Example

<b>Operating Svstsm</b>		
Redhat release	6.2	7.1
Kernel release	2.2.20	2.4.3
Kernel source	RedHat.Com	redhat.com
<b>shells</b>		
tcsh	4.5.	6.09
zsh	8.2	not tested
bash	not tested	2.03
Login scripts	HEPiX vers x	FUE login 1.6

<b>Databases. Data Access</b>		
ORACLE	8.1	7.1.6, 7.3.4, 8.1.5
Obiectivity	7.3	not tested
mysal	Not used at CERN	3.23.41

# Software Certification

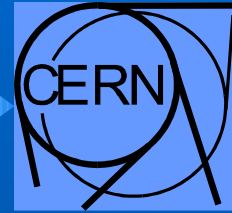
## – The Challenge



- Too many modules to certify, too often, many sites repeating the same or similar work
- Grid computing implies the need for standard environments across sites – for example the CMS experiment will expect to find a given Linux version, given compiler version, libraries, etc. at all their Tier sites
- Many sites support multiple experiments, which means supporting multiple environments in parallel



# Software Certification Network

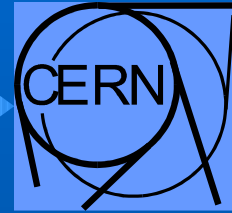


We propose to create a "network" of people at the different sites who are directly concerned with software certification, in particular with Linux and Solaris.

These individuals should discuss matters of mutual concern such as

- what modules need to be certified, in particular how to decide what modules are needed for specific user communities. This is in relation in particular to Grid activities where experiments will request specific environments across multiple sites
- how certification is carried out
- which versions to certify
- Any commonalities? Any chance of unification?

# Software Certification Network

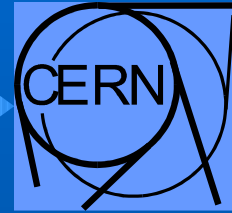


## Timetable

- HEPiX Catania - expose the proposal
- End April - those sites which wish to participate should nominate an individual (to Alan Silverman or Jan Iven)
- Early May - creation of a mailing list for a first exchange of ideas on a better mandate and the first discussions
- Early June - a first (and possibly only) face-to-face meeting to get the process started

It is imagined that most of the discussions will be electronic and few physical meetings will be necessary although the idea of special meetings around HEPiX and GGF meetings should be kept open.

# Large Computer Centre Operations

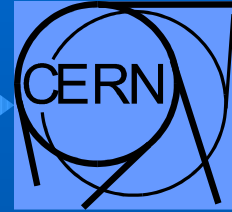


- Surveyed the major sites (BNL, Caltech, CERN, FNAL, RAL, SLAC)
- Do you have operators round the clock?
- Do you or have you experimented with “lights out”?
- Do you have procedures for emergency shutdown, and restart, of your Centre?
- Have you ever tested these?
- Similar questions on use of UPS and if sites have and test power cycle procedures

<b>SITE</b>	<b>Operator Cover</b>	<b>Lights-out Operation</b>	<b>Monitoring and Alarm Tools</b>
FNAL	24 x 7 In-house team	Under active consideration - question of possibly eliminating the night cover	NGOP for monitoring and alarms to the Operators; use of Remedy for trouble tickets
SLAC	None	Fully lights out	Remote system admin; monitoring with automatic alarms
BNL	24 x 7 during runs; day shifts via operators, overnight via the experiments	No overnight cover when no beam	
Jefferson	None	Fully lights out	Use of tape robots plus monitoring via the mon tool with alarms to an on-call sys admin
RAL	8 x 5 in-house team plus callout	Outside prime shift	Monitoring with alarms
IN2P3	70% cover by in-house team	Overnight only	Monitoring with alarms to the Operators
Caltech	Staff by day, Students at nights and weekends	Has been considered but rejected - centre too complex, equipment too special	Very little automation - budget concerns
NERSC	24 x 7 In-house team	Considered but no easy, cheap solution found	Monitoring using netsaint and snmp with alarms
DESY Zuethen	Day shift cover, Mon to Frid	None	Monitoring with alarms
DESY Hamburg	Cover during day shift only		Operator on call out overnight
CERN	24 x 7 Outsource contract	Considered but no easy, cheap solution found	CNSURE for monitoring and alarms to the Operators, Remedy trouble tickets to sys admin team; use of tape robots

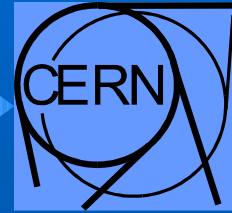
SITE	UPS Protection	UPS Test	Power Cycle Test	Shutdown Procedure	Shutdown Procedure Test	Emergency Shutdown Procedure	Emergency Shutdown Procedure Test	Startup Procedure	Startup Procedure Test
FNAL	Yes	Monthly	Yes	Yes, by owner group	Twice per year	Not needed, UPS fed by diesel	N/A	Semi-formal	Yes
SLAC	Yes	Monthly	Not planned but yes!	Yes	3-4 times per year	Yes	No, rely on full shutdown procedure	Yes	3-4 times per year
BNL									
Jefferson	Yes; also a backup generator	UPS annually, generator weekly	By accident about once per year	Yes	Yes	Yes	Yes	Yes	Yes and kept up-to-date
RAL	For some servers	No	No	Yes	Not for some time	No	N/A	Yes	Not for some time
IN2P3									
Caltech	For some services	Twice per year	No	Yes	Not for some time	No, rely on UPS being refuelled	N/A	Think yes but not certain	Not for some time
NERSC	For some servers	No	No	Yes for the large servers	No		During last power cut	Think yes but not certain	Only when needed
DESY Zuethen									
DESY Hamburg									
CERN	Yes	Occasional	Occasional	Recently created	and tested	Working on this	Not yet	Recently created	and tested

# OpenAFS



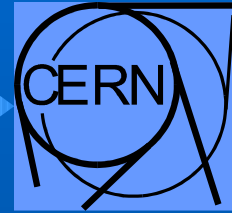
- HEPiX at JLab – IBM warned us they would be dropping support of AFS
- HEPiX Paris and NERSC – use of OpenAFS grows in HEP
- HEPiX Catania – IBM fulfills a promise!
- All (almost) HEP sites depend on (open)AFS
- Can we leave its future to the wider OpenAFS community?
- Should we form an action group inside or in parallel to the OpenAFS community?
- Should we talk to a vendor about “support” (SLAC model)
- Is someone willing to drive this, gather input from sites?

# Collaborations



- Many pleas this week – Andrei's Storage Lab, Cary's storage benchmarks
- Good idea but make it easy – package the tests, results gathering, etc; commit to publish results

# Plans



- Re-orient the Software Certification effort, promote interested parties to work in parallel if possible, targetting this work at real usage for the Grids
- Work on the Cluster Guide and hopefully issue a zero'th release before the next HEPiX
- Work with Grid and HEPiX colleagues to instill the idea of using HEPiX for technical discussions related to using our Centres in Grids
- Organise a 2 (or 3) day workshop on Large Clusters at the next HEPiX
- Offer a forum for OpenAFS collaboration