# UPGRADES TO THE INFRASTRUCTURE AND MANAGEMENT OF THE OPERATOR WORKSTATIONS AND SERVERS FOR RUN 2 OF THE CERN ACCELERATOR COMPLEX

A. Bland, S.T. Page, CERN, Geneva, Switzerland

*Abstract*

The Controls Group of the CERN Beams Department provides more than 400 operator workstations in the CERN Control Centre (CCC) and technical buildings of the accelerators, plus 300 servers in the server room (CCR) of the CCC. During the long shutdown of the accelerators that started in February 2013, many upgrades were done to improve this infrastructure in view of the higher-energy LHC run. The Engineering Department improved the electrical supply with fully redundant UPS, on-site diesel generators and for the CCR, water and air cooling systems. The Information Technology Department increased network bandwidth for the servers by a factor of 10 and introduced a pilot multicast service for the video streaming of the accelerator status displays and beam cameras. The Controls Group removed dependencies on network file systems for the operator accounts they manage for the Linacs, Booster, PS, ISOLDE, AD, CTF3, SPS, LHC and cryogenics. It also moved away from system administration based on shell scripts to using modern tools like version-controlled Ansible playbooks, which are now used for installation, day-to-day re-configuration and staged updates during technical stops.

## INTRODUCTION

The Controls Group of the Beams Department at CERN manage disk-based Linux workstations and servers for the operation of all of CERN's accelerators, cryogenics and technical infrastructure.

The workstations are business-class Personal Computer (PC) desktops ranging from 2006 to 2015 models running Scientific Linux CERN 6 (SLC6, based on Red Hat Enterprise Linux 6). Eighty of them are concentrated in the CERN Control Centre (CCC) on the Prévessin site in France and the rest are distributed across more than 100 technical buildings including underground areas of the Large Hadron Collider (LHC). All workstations can run the software of any part of the accelerator complex; restrictions on what can be done from where are enforced by the Role-Based Access Control system.

The servers are almost entirely in the CERN Control Centre server room (CCR) with a small number two kilometres away on the Meyrin site in Switzerland for redundancy and recovery purposes. The majority were upgraded to SLC6, only ten still run SLC5.

Although the accelerators themselves have well-defined year-end technical stops and around 2 or 3 technical stops during the year for upgrades, the technical infrastructure operation (electricity, cooling and ventilation and safety installations) and cryogenic parts of the control system must run 24 hours a day, 365 days per year.

## TIMELINE

Most of the upgrades were performed between the start of Long Shutdown 1 (LS1) [1] in February 2013 and the first LHC circulating beams of Run 2 in April 2015, see Table 1. Not all of this time was suitable for infrastructure upgrades as the injectors for the LHC, but not the LHC itself, ran from April 2014 to December 2014.

Table 1: Principal Dates

| Date | Activity |
|---|---|
| Q1 2006 | First accelerator operation from the CCC |
| Q3 2008 | LHC first circulating beams |
| Q1 2010 | LHC first physics at 7 TeV |
| Q1 2013 | Long Shutdown 1 starts |
| Q2 2014 | Injector startup |
| Q2 2015 | LHC Run 2: first circulating beams and physics at 13 TeV |
| Q1 2019 | Long Shutdown 2 expected, 18 months duration [2] |

Another complication for upgrades between the start of LS1 and April 2014 was that not all accelerators stopped; the Compact Linear Collider Test Facility (CTF3) and Linear Accelerator 4 (Linac4) tests had to continue, plus the LHC Magnet Test Bench (MTB) continued operation.

## OPERATIONS REDUNDANCY

The Technical Infrastructure (TI) operator environment has been replicated in the fire station on the Meyrin site with four workstations, each with two screens. The main TI telephone number can be switched to this installation. One TI shift per month is run from the fire station to ensure that all activities are possible from there.

Similarly the firemen's environment in the fire station has been replicated in the CCC with five desktops, the Terrestrial Trunked Radio (TETRA) system and the hardwired fire alarm panel. One shift per week is run from the CCC to test the installation.

## ELECTRICAL NETWORK

From first operation of the CCC in 2006 up to LS1, powering was already possible either from the Swiss or French national electrical network. Diesel backup if needed came from the Meyrin site and three sets of Uninterruptible Power Supplies (UPS) were locally installed on the Prévessin site providing N+1 redundancy.



ISBN 978-3-95450-148-9

However, there was only one power line from the UPS into the CCC. The server room cooling and ventilation could not be powered from diesels or UPS which implied that when running from these sources, many of the servers would need to be switched off within half an hour to prevent overheating. Workstations only had one UPS-supplied power source which would be shed after 15 minutes. Servers had two UPS power sources, one of which would be load shed after 15 minutes and the other when the UPS batteries were empty.

A major project by the Electrical Group of the Engineering Department (EN/EL) during LS1 [3] made the following improvements:

- Installed three diesel generators on the Prévessin site providing N+1 redundancy.
- Constructed a new electrical building divided in two with a two-hour concrete firewall providing two sets of UPS (2N+1) and associated switchgear (2N).
- Two power cable routes into the CCC, separated to the greatest extent possible.
- Two completely independent sources of power to all essential equipment in the CCC and the CCR racks.
- Connected critical equipment with only one power input such as workstations for technical infrastructure, cryogenics and access to auto transfer switches which seamlessly allow operation to continue in the event of a supply path failure.
- Consequently, maintenance of the electrical system can be performed while continuing service.

No down time of the rack-mount and blade servers occurred during this upgrade as they have at least dual power supplies, so it was possible to remove one of the old power inputs then add the first new one, remove the second old input and finally add the second new one.

It is to be noted that the electrical supervision system runs on servers in the CCR.

## COOLING AND VENTILATION

The pre-LS1 cooling and ventilation of the CCR used the chilled water supply coming from the neighbouring Super Proton Synchrotron (SPS) technical building, so maintenance in that building, power cuts, emergency stop tests and failures led to overheating of the servers. The amount of air that could be ventilated out of the building was very small, so even when cold outside, the system administrators would have to start a portable fan and open the doors.

A major project by the Cooling and Ventilation Group of the Engineering Department (EN/CV) during LS1 [4] made the following improvements:

- Two separate cooling and ventilation systems in CCR. The cooling and ventilation systems of the critical CCR and telecom (network) rooms have been separated from the non-critical system for CCC operators and meeting rooms.
- Cooling power supply is backed up by a diesel generator and UPS power.
- Redundant power supply for the cooling system of the CCR and the telecom room which can be powered from either of the two CCR power sources.
- Added dedicated chillers for the CCR and telecom room with redundant two-out-of-three chiller configuration where there is a chiller on stand-by.
- Water-cooled rack doors for HP ProLiant blades remove the concentrated heat in the server area.
- Dedicated cold air entry in front of server racks and hot air removal behind.
- Consequently, maintenance can be performed while continuing service.

Due to the extra depth of the water-cooled rack doors, the central row of racks in the CCR had to be moved by 30 cm and this was performed with no cut in the service. Around 20 servers not needed during LS1 were stopped to reduce the heat load on the two portable fans used while the cooling system was being upgraded. No other down time occurred.

It is to be noted that the cooling and ventilation supervision system runs on servers in the CCR.

## NETWORK

The Communication Systems Group of the Information Technology Department (IT/CS) installs and manages essentially all of the Ethernet outlets, cables, fibres, switches and routers needed for accelerator operation. A minor exception is that typically each HP blade enclosure contains two 10 gigabit/second switches which are the responsibility of the Controls Group.

During LS1 IT/CS:

- Increased the number of Ethernet outlets by 25% in the CCC and the CCR.
- Re-cabled and increased the number of patch panels between the switches and the outlets.
- For the CCC, upgraded older HP ProCurve 3400 switches with a 1 gigabit/second copper uplink to the 3500 model with optional 10 gigabit/second uplink.
- Installed one HP ProCurve 8206 router in each of the three rows of server racks.
- Connected 10 gigabit/second downlinks from the new routers to each of our blade switches.
- Installed redundant routers in parallel with both the main CCC/CCR router and the IT Computer Centre router for the accelerator network. Now maintenance can be performed on one of each of these sets of routers while continuing service without the typical 10 minute down time previously experienced during firmware upgrades.
- Enabled multicast routing in the CCC/CCR accelerator network router as a pilot service. These "targeted broadcasts" allow efficient video streaming of the accelerator status displays and beam cameras to any workstation in the CCC as the network load between the router and switches of the sender and client does not increase with the number of clients - streams are only sent once. In case of problems with the pilot service, video streams will fall back to the pre-LS1 use



ISBN 978-3-95450-148-9

of our Video on Demand service via a central server, at the cost of higher network bandwidth.

During the CCC re-cabling, patch panel and switch upgrades, the TI operators moved to their backup location and the small amount of cryogenic operation occurring at that time was performed from the Cryogenics local control rooms at the LHC points.

The move of the servers from the main CCC/CCR router to the three new HP ProCurve 8206 routers gave only 10 seconds of network loss for each server.

The redundant router installations needed a restart of each of the switches under them, typically provoking around 90 seconds of network loss for all connected hosts.

## WORKSTATIONS

From 2010 onwards, the CCC workstations were business-class HP Elite 8100 PCs (Intel Core i7-860 2.8 GHz processor, Lynnfield generation, 8 cores + threads total, 8 GB RAM, 250 GB hard disk). This had the optimal processor performance for the time, but did not have any on-board graphics processor, so all of them had either one or two Nvidia Quadro NVS 295 business graphics cards to drive the one to four monitors of a typical CCC workstation.

In 2015, a joint specification with the IT Department for office PCs for CERN chose the Optiplex 9020 Mini Tower Business-class Desktop from Dell. The specific requirement for controls use was the support of Intel Active Management Technology (AMT) [5] which permits remote restart and out-of-band BIOS management of PCs operating below the Operating System (OS) layer with no extra Ethernet connection; this saves long trips to the LHC underground areas simply to power back on a workstation for an update.

The chosen model has an Intel Core i7-4790 3.6 GHz processor, Haswell Refresh generation, 8 cores + threads total, 16 GB RAM and 256 GB Solid-State Drive (SSD). As 3 monitors can be driven with the on-board Intel graphics, 90% of the CCC workstations need no external graphics cards to be added, reducing heat load, spares requirements and cost. At the same time as the 110 CCC workstations (including 30 running Windows 7) were upgraded, the 285 monitors were also renewed. The recuperated Elite 8100 PCs, being already AMT capable, have been reused to upgrade the non-remote manageable workstations dating from 2006 in the technical buildings and LHC underground areas.

### *Accounts*

Around 1000 user accounts have access to the accelerator control system, typically with home directories on Andrew File System (AFS) servers managed by the IT Department or on Network File System (NFS) servers managed by the Controls Group in the CCR.

For Run 2 of the accelerator complex, the operator accounts have local home directories physically on the hard drive or SSD of the workstation. This gives isolation from difficulties accessing the central NFS infrastructure and reduces network load. In rooms such as the CCC with access control, the workstation performs an auto-login to the desired account. The login starts the Java Common Console Manager which provides a per-Operations-team menu of programs that can be run, defined in the Controls Configuration Database (CCDB).

## SERVERS

Most of the HP DL380 G5 rack-mounted servers installed since the startup of the CCC in 2006 have been retired. The replacements are currently HP ProLiant BL460c Gen9 half-height blade servers (two Intel Xeon E5-2630 v3 2.4 GHz processors, Haswell generation, 32 cores + threads total, 64 GB RAM, two 600 GB SAS hard disks in a RAID 1 array), 16 of which can be installed in an HP BladeSystem c7000 enclosure. Systems needing a large number of disks such as NFS servers and boot servers have been upgraded with rack-mounted DL380 Gen9 hardware (two Intel Xeon E5-2637 v3 3.5 GHz processors, 16 cores + threads total). The server usage is shown in Table 2.

Table 2: Server Usage

| No. | Hardware | Usage |
|---|---|---|
| 80 | Blade | Beam control and technical infrastructure based on Java |
| 130 | Blade | Industrial controls based on Siemens WinCC OA |
| 5 | Blade | Testing and measurement using National Instruments LabVIEW |
| 10 | Rack-mount | Network File System servers |
| 5 | Rack-mount | Boot servers for the 1500 diskless front ends |
| 8 | Blade | LHC real-time orbit feedback and software interlocks |
| 12 | All types | Spares and test bench |
| 20 | Blade | Development, web, samba access from Windows, interactive login etc. |
| 10 | Desktop | Media centres for CCC wall displays |
| 20 | Desktop | Accelerator status displays |

CERN General Machine Timing PCI receiver cards are needed in servers for two timing-critical areas:

- Proton Synchrotron (PS), SPS and LHC software interlock system (4 servers).
- LHC real-time orbit feedback (4 servers).

Early in LS1, these were upgraded from rack-mounted systems to HP Gen8 Blades with the PCI timing receiver cards installed in HP BladeSystem PCI Expansion Blades.



ISBN 978-3-95450-148-9

## SYSTEM ADMINISTRATION

Until LS1, a per-OS shell script was run after the base OS was installed to configure a workstation ready for an operator to use it or to run Java processes in the case of a typical server. The same per-OS shell script was also run early in the morning of every working day to keep the configuration up-to-date.

During LS1, a major investment was made by the system administrators to switch these per-OS shell scripts to the Ansible configuration management system [6].

The host name of every machine we manage is defined in a group in the Ansible inventory file. In our case, we generally map groups of machines to roles such as workstation, media centre, server, WinCC Open Architecture server, NFS server etc.

Ansible playbooks, written in YAML (a human-readable data format), code the specific idempotent (repeatable with the same result) tasks needed to configure the characteristics of a role. Ansible has built-in modules for installing software packages, configuring services, scheduling tasks and adding, removing or modifying lines in configuration files. Any functionality not implemented as a module can be coded with a line of shell script or a custom module can be developed, typically in the Python language. Tasks can have a "when" condition permitting them to be skipped if not applicable. This has allowed us to have one Ansible playbook that manages all the operating systems we use: RHEL5, SLC5, SLC6 and CentOS 7. This means that a new requirement for the control system is added by default to all operating systems, but can, if needed, be coded differently for each one. Any errors in a playbook stop execution, but they can be ignored if required. The end result of the use of Ansible is that system administration becomes more structured and less subject to the individual shell script author's coding style than before.

Ansible's more traditional usage is for a system administrator on a central management host to push out updates to potentially hundreds of machines in parallel. In two hours during our one-day technical stops, we have been able to update all 400 workstations and 100 servers from SLC6.6 to SLC6.7, which typically takes about half an hour per machine including the final reboot.

Ansible, being agentless and using OpenSSH for remote node management, was found to be the best programmable infrastructure choice for our environment.

## VERSION CONTROL

Before LS1, the configuration files and scripts needed for the management of the workstations and servers were either using file-based Revision Control System (RCS) or simple dated backup copies, accessed live over NFS. If more than one system administrator was modifying configurations at the same time, there was the possibility of conflicting changes.

During LS1, the key configuration files and in particular the daily Ansible playbook, have been put into a Git [7] version control repository. These files are updated from the Git server via the network onto each machine every day before the daily Ansible run.

System administrators working on a new feature for the control system can make a "branch" in the Git repository which they can safely check on a chosen test bench machine without affecting the production environment then later “merge” into production. All configuration changes are traceable to the author and their intended action.

## CONCLUSION

A massive upgrade campaign with a strong focus on redundancy, reliability and efficiency of operation and minimal downtime during implementation was completed during Long Shutdown 1 to achieve the requirements of Run 2 of the LHC.

The powering, cooling and network bandwidth upgrades performed during LS1 allow a considerable expansion of the number of servers during Run 2 and the system administration based on Ansible and Git will scale to the needs.

With the Electrical and Cooling and Ventilation Groups there has been an excellent symbiosis, helped by the fact that their supervision system is running on the servers managed by the Controls Group in the CCC server room.

## ACKNOWLEDGEMENTS

Staff and contractors of the EN/EL, EN/CV and IT/CS Groups for the careful planning and implementation of the upgrades.

The Beams Department Operation Group for their patience during the disturbance.

Mikhail Grozak working as a CERN Project Associate wrote the initial Ansible implementation.

## REFERENCES

[1] K. Foraz et al., “LS1 First Long Shutdown of LHC and its Injector Chains”, 5th International Particle Accelerator Conference, Dresden, Germany, pp. 1316 (2014).

[2] M. Bernardini, K. Foraz, “Long Shutdown 2 @ LHC”, LHC Performance Workshop, Chamonix, France, pp. 290-293 (2014).

[3] F. Duval, “LS1: electrical engineering upgrades and consolidation”, Chamonix 2012 Workshop on LHC Performance, Chamonix, France, pp. 248-251 (2012).

[4] M. Nonis, “EN-CV during LS1: upgrade, consolidation, maintenance, operation”, Chamonix 2012 Workshop on LHC Performance, Chamonix, France, pp. 252 (2012).

[5] AMT website: http://www.intel.com/technology/platform-technology/intel-amt/

[6] Ansible website: http://www.ansible.com/

[7] Git website: http://git-scm.com/

ISBN 978-3-95450-148-9