Systems Management of the DAQ systems

M. DOBSON
ON BEHALF OF
ALL EXPERIMENTS

Special thanks to Diana, Sergio, Ulrich, Loïc, Niko

Purpose of System Administration

- Keep the DAQ and sub-detector systems in the best possible state to take data !
- Help and contribute to the design of the DAQ systems
- Large farms and network, including supporting HW (NAS, etc…) need maximum up time (high availability)
  - Minimize single points of failure: redundant systems
  - Good monitoring for fast diagnostics
  - Fast recovery (configuration management, local installs or netboot)
  - Adapting to loss of HW: virtualization, HA tools (corosync, pacemaker, HA proxies)
  - Redundant networks and connections
  - Live with GPN disconnect (local data storage)
- Run Efficiency and resilience
  - Basically identical to last DAQ@LHC forum (see presentations at http://indico.cern.ch/event/217480/)

Content

- Run2 System Sizes
- Operating Systems
- Installation and booting
- Configuration management
  - Package Repositories
- Virtualization
- Monitoring
  - Network monitoring
- Support
- Hardware, procurement, and maintenance
- New HW challenges: embedded Linux, SoC, ATCA/uTCA

Run 2 System sizes

- ALICE
  - 155 readout machines, 20 event builders, 20 services servers, 18 switches (data + control)
- ATLAS
  - 3600 machines (2500 netboot), ~200 control switches, 75 data switches, 480 HLT nodes being delivered, 130 DCS nodes (Linux)
- CMS:
  - 1250 PC DAQ related (including farm), 200 sub-detectors, 100 DCS, 50 central, 100 / 70 control / data switches
- LHCb
  - 1750 farm nodes, 100 Servers, 300 VMs, 200 switches

Operating Systems

- Currently based on one of the latest SLC6 releases
- LHCb has a few central servers on CERN CentOS 7:
  - New control room machines
  - Web servers being migrated now

- Migration to CERN CentOS 7:
  - CMS plans to start with hypervisors and some central services during Q3-Q4 2016, DAQ tests in Q4 2016 for migration in YETS 2016 (new DAQ SW release only on CC7)
  - LHCb: no firm plans for DAQ or farm yet
  - ATLAS: will start looking at it, might need it for next version of WinCC at next YETS
  - ALICE: planned for some services (monitoring/shared file systems)

Installation/booting of nodes

- Network booting:
  - LHCb and ATLAS have many/most nodes network booted
    - LHCb: control room machines, farm nodes, credit card PCs
    - Infrastructure for NFS OS mounts and boot servers
    - Hierarchical structure
- Local installation
  - CMS and ALICE have locally installed nodes
  - LHCb have some locally installed nodes (other servers)
  - ATLAS has DCS, infrastructure servers, and also DAQ infrastructure
- Centralized storage
  - ATLAS and CMS have NetApp NAS for home directories and project areas
  - LHCb: DDN for physics data, home/group/project directories. NetApp for the virtualization infrastructure
  - ALICE: shared file system for the control room machines, SAN file system (1PB) as buffer for the event builders
- Virtualization: see later

Configuration Management

- Puppet: ATLAS preceded IT, CMS and LHCb followed IT
  - Versions:
    - ATLAS version 3.3.2 & migrated to Puppet DB
    - CMS version 2.X, planning migration to v3 by summer 2016
    - LHCb version 3.5, ideas to go to version 4
    - ALICE version 3.8.6
  - Install:
    - ALICE: basic kickstart, then puppet
    - ATLAS: uses own ConfDB for provisioning (no plans for Foreman)
    - CMS: Foreman used to kickstart
  - All use pull mechanism
    - CMS & LHCb: puppet agent, respectively 30m, 2h (with splay), LHCb also on netboot
    - ATLAS: Puppet used for netboot and localboot
    - ATLAS uses Puppet also for netboot image creation and boot time specialization
      - Puppet apply for netboot nodes via cron job every hour
- Plans are to continue with Puppet

Package repositories / software distribution

- Software repositories (OS, core):
  - Regular mirroring of the IT yum repositories
    - Implemented as dedicated snapshots
      - Able to go back in time
      - Versioned test/production/… for ATLAS
    - ATLAS/ALICE have dedicated security repo to only bring in security updates (not general ones)
    - LHCb use BTRFS features of versioning (snapshoting)
    - ALICE: snapshot ~once per year
    - ATLAS & CMS: use hard links for duplicate files
- DAQ & sub-system software
  - ATLAS: distributed hierarchically by file servers (rsync + NFS)
  - ALICE & CMS: use RPMs, and software repositories. CMS has a Dropbox built on top.
  - LHCb: use CVMFS

Virtualisation

More and more use as indicated at last workshop

- ATLAS:
  - 6 Gateways
  - 2 detector nodes
  - 4 Domain controllers (IT)
  - DCS (~ 40 Windows VMs, planned migration to Linux)
  - LDAP servers (9)
  - DAQ web service backend (10)
  - Technical infra (SLIMOS) (2)
  - 3 public nodes
- LHCb:
  - Login services
  - Infrastructure services (some)
  - Most DCS servers (iSCSI booted CCPC for HW access)
- CMS:
  - Domain controllers (IT)
  - Gateways
  - Infrastructure services (some)
  - Detector machines
  - Some DCS (windows VMs)
  - DAQ services (run control)
- ALICE:
  - Gateway services (10 VMs per server)
  - Critical services (1 VM per server)

Virtualization 2

- Technologies:
  - ATLAS use KVM (Kernel-based Virtual Machine) hypervisor
  - CMS use oVirt clusters with underlying KVM
  - LHCb use Red Hat Enterprise Virtualization (RHEV) based on oVirt and KVM
  - ALICE use HyperV (WinServer 2012R2), also snapshots
- Live migrations
  - ALICE, CMS & LHCb: yes
  - ATLAS: no (no suitable image storage provisioned, conscious decision, spread risk on more servers), could be reviewed in CC7 as no need for common storage
- Migration on failures of HW ?
  - CMS: HA feature of oVirt
  - ALICE: fail over to other hypervisors
  - LHCb: HA feature of RHEV
  - ATLAS: restart on different Hypervisor from image backup
- Alternative usage of HLT Farms:
  - Cloud usage (ATLAS, CMS): Openstack based, VMs prepared by offline teams
  - LHCb run Dirac SW for offline processing during shutdowns (no cloud)

Monitoring

- Large infrastructure must be monitored automatically
  - Proactively warn of failure or degradation in system
  - Avoid or minimize downtime
- What is monitoring ?
  - Data collection
  - Visualization (performance, health)
  - Alerting (SMS, email)
- Most experiments use Icinga2
  - Gearman/mod_gearman (queue system) deprecated

Monitoring 2

- ATLAS
  - Ganglia for performance data
  - Icinga2 (gets some data from Ganglia)
  - Icinga config generated from ConfDB
  - 70k checks
  - Icinga2Web
  - scripts for massive execution
  - notifications being improved for wider audience
- LHCb
  - Icinga2
  - Configuration managed by Puppet using info from Foreman
- CMS
  - Ganglia for some performance data
  - Icinga2 (manual config)
  - Icinga2Web
- ALICE
  - Zabbix
  - No more updates in SLC6 for the server part
  - Migration of servers to CC7

Control Network: config & monitoring

- ALICE:
  - Installed and managed by ALICE
  - SNMP traps for the monitoring
  - Static configs, tftp config load on boot under study
- CMS:
  - Control network configured & monitored by IT
  - Spectrum available to us.
  - Icinga2 monitors switches being up/down and sets dependencies
- ATLAS:
  - Part of control network managed by DAQ network team
  - IT configure and manage the rest (Spectrum available, Icinga monitored also)
  - Icinga (version1) for device/link health monitoring and network traffic alerts
  - Netis for device traffic monitoring and device environmental metrics.
- LHCb:
  - Installed & managed by DAQ
  - Cacti and Icinga monitoring

Support

- Ticket systems used to track issues or requests
  - ALICE & CMS use JIRA (IT provided)
  - ATLAS uses Redmine (local, started before IT JIRA available)
  - LHCb uses ServiceNow (IT provided)
- Urgent matters are managed via on-call teams with different philosophies
  - ALICE: DAQ on-call as first line, dispatches other experts as needed
  - CMS & LHCb: DAQ on-call is the first line, then SysAdmins
  - ATLAS: direct call to TDAQ SysAdmins

HW, Procurement & maintenance

- Do experiments follow IT tenders ? For what HW ? How does maintenance change?
  - LHCb:
    - Try to follow IT tenders whenever possible
    - No difference as they have always done the maintenance themselves
  - ALICE do not follow IT tender for the server HW (due to RORC HW specifics), however uses market survey
    - 5 year on-site warranty, only small repairs done by SysAdmins (e.g. disk in holder)
  - ATLAS follow IT tenders
    - Additional communication layer (IT), longer part replacement
    - More issues seen than on previous (non IT) tenders
  - CMS follows IT tenders for farms
    - Maintenance is radically different, before had 5-year on-site warranty

- HW inventory, what do we do?:
  - HW history and issue tracking: Redmine and JIRA not well suited
  - IT tools very integrated in their custom workflow
  - CMS have used OCS inventory (open source technical management solution of IT assets) and GLPI (Information Resource-Manager with an Administration-Interface). It is being revived. Collaboration between experiments is probably good here.

New HW challenges

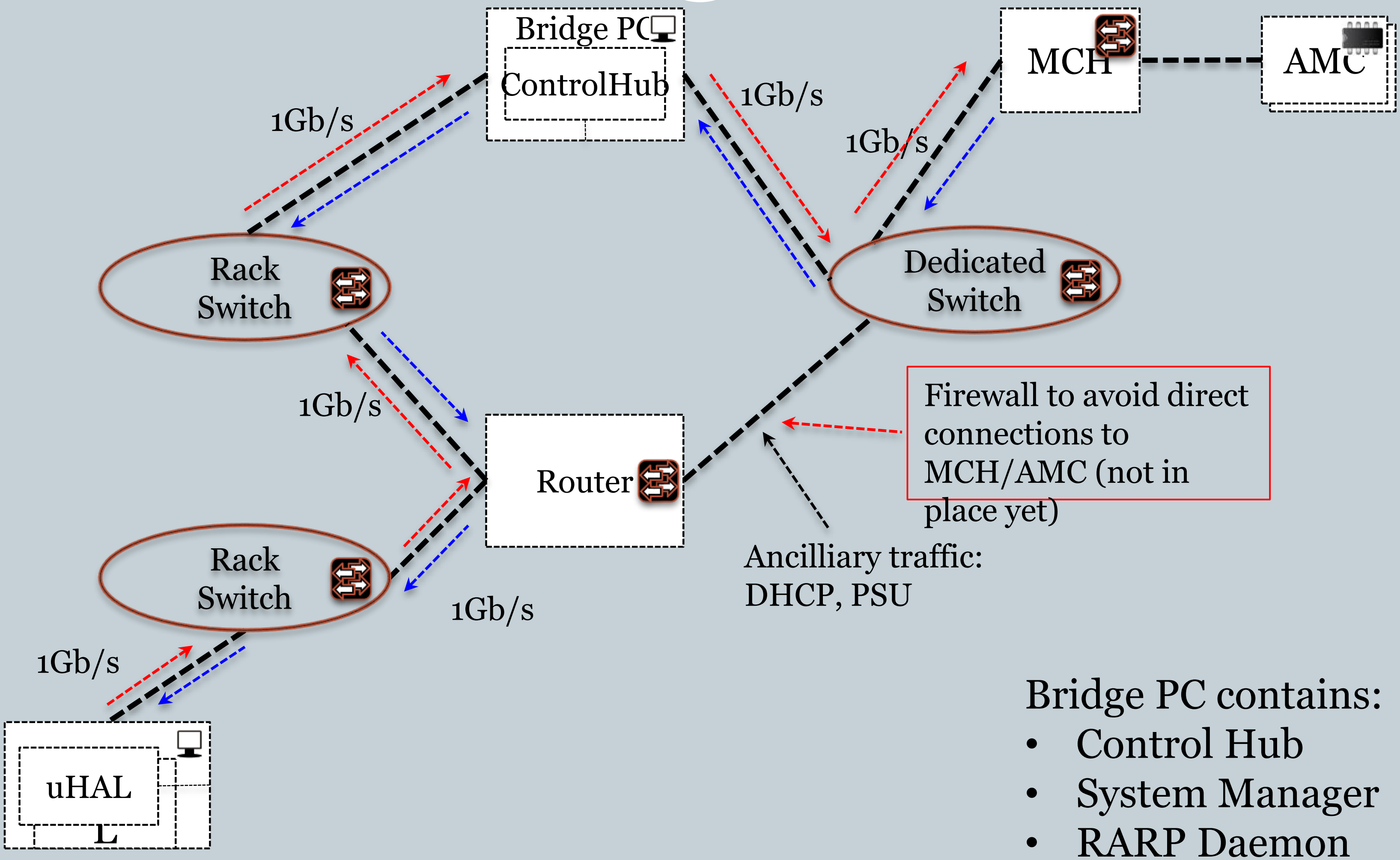
- Embedded Linux, SoC
  - ATLAS: 2 sub-detectors started using embedded Linux
    - Security documents required for the management of the security updates by them
  - LHCb: Credit card PCs (Atom based), standard pinout, not really SoC
  - First few Raspberry Pi devices, some Arduino (controllers)
    - How do you manage them, also for security updates etc… ?
- ATCA and uTCA hardware
  - Has needed much prototyping and testing
  - ATLAS: 5 sub-detector using ATCA
    - Different manufacturers adopted (Asis, Pentair, Schroff), Pigeon Point for the shelf managers
  - CMS: 6-7 sub-detectors using uTCA
    - Different manufacturers used for MCHs (NAT, Vadatech), and crates (Schroff, Vadatech), specific backplanes for certain lines (TTC distribution)
  - ALICE and LHCb: happily xTCA free !

HW Challenges: uTCA/ATCA

- CMS is uTCA based: 6U chassis with 12 AMCs + MCH + data concentrator (AMC13)
  - Ethernet to the MCH (control/monitoring of crate)
  - Using mainly IPBus to talk over Ethernet (1Gb) to AMCs (slow control, monitoring and local readout)
    - Has many implications (see next slide) as endpoints are simple
  - Data paths are through backplane to AMC13 mainly + readout from there
  - Some people use PCIe bridges on MCH to make the crate look like an extension of the controlling PC PCIe bus (point to point links with single points of failure)
  - CMS will likely go to ATCA for Run 3 (more real estate for the electronics)
  - Some people have a SoC on the AMC board with the FPGA (Zynq by Xilinx) running some embedded version of Linux

- ATLAS use ATCA
  - Switch fabric inside crate used + additional Switch cards for external connectivity
  - IP addresses allocated via DHCP, some hardcoded, IPBus IP allocated via i2c bus
  - IPBus used for configuration and update
  - Shelf manager provides SNMP access for DCS

uTCA in CMS



Bridge PC contains:

- Control Hub
- System Manager
- RARP Daemon

Conclusion

- DAQ clusters are no longer exceptionally large
  - Can "follow" industry development and adopt "standard tools" (e.g. Puppet, Icinga2)
  - However variety of HW and uptime requirements are higher
  - Workload per host higher than most IT, grid farms, virtualized clusters

- DAQ is mainly NOT virtualized
  - Squeeze most performance and lowest latency from COTS HW
  - Dedicated data network connections
  - This has much impact:
    - On the overall architecture
    - On SysAdmin load (harder than fully virtualized environment)

- Standard IT technologies going further towards detectors
  - More versatile clients for SysAdmins.
  - New technologies (SOC, embedded Linux) with their security implications

➔SysAdmins should be an integral part of designing RUN3/4 DAQ/dataflow systems

- Much can be shared between experiments (and IT)
  - Knowledge, expertise
  - Investigations, research, experience
  - Restart X-experiment meetings