Procurement trends at CERN

Eric Bonfillou

Miguel Santos

Olof Bärring

Outline

- Intro to Procurement of servers & storage to CERN computer center
- Interesting stuff in 2011
  - ‘Disk server’ evolution
  - Solid state drives
  - On-site repairs
  - Hard disk headaches
- Foreseen changes in 2012 and beyond
- Conclusions

CERN/IT Procurement team

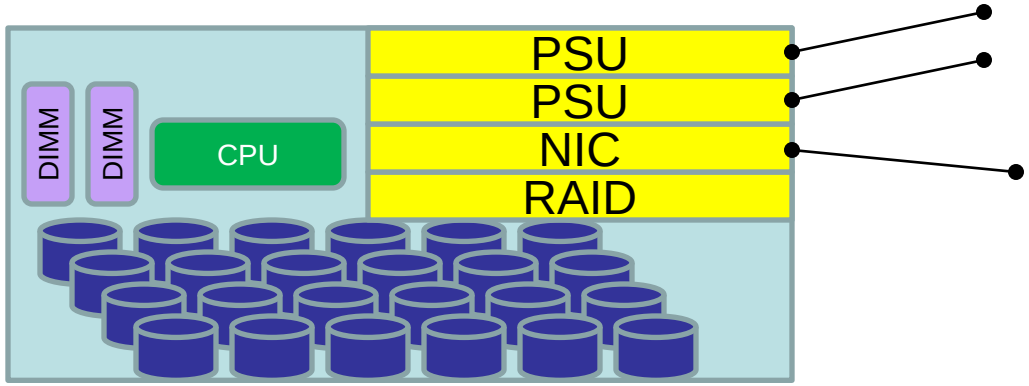
- Technical responsible for hardware procurement to CERN IT computer center:
  - Servers & storage for physics (bulk capacity)
  - Servers & storage for most infrastructure services
- Tasks
  - Tender specs
  - Bids evaluation and sample testing
  - Burn-in and acceptance tests
  - Assist (service mgrs, sysadmins) in analysis and problem resolution of hardware issues
  - Maintain *technical* contacts with suppliers and manufacturers
  - Develop and maintain tools for hardware monitoring and management tools
  - Plan installations with CC operations teams
  - Follow trends and perform R&D

‘Disk server’ evolution (trade-in the hw RAID)

CERN IT Department

- Until recently: large direct attached storage (DAS) servers
  - Single socket (not CPU bound)
  - Small memory (not memory bound)
  - 10GbE
  - RAID card
  - 24 or 36 internal disks attached to a backplane or multi-lane SATA cables
- Worked wonderfully as long as disks <1TB
- Above 1TB hw RAID becomes inadequate
  - Large disks→ large RAID partitions → long & heavy rebuilds upon HDD failures. Real risk for 2nd failure while rebuilding
  - Write cache means BBU and associated failures (super capacitors better but expensive)
  - Not hot-swappable and they do fail (in particular BBUs)
  - Firmware spaghetti (BIOS – controller – disks)
  - Maintain expertise with clunky proprietary tools
- All-in-a-box → 50TB offline when swapping a DIMM

Old ‘disk server’

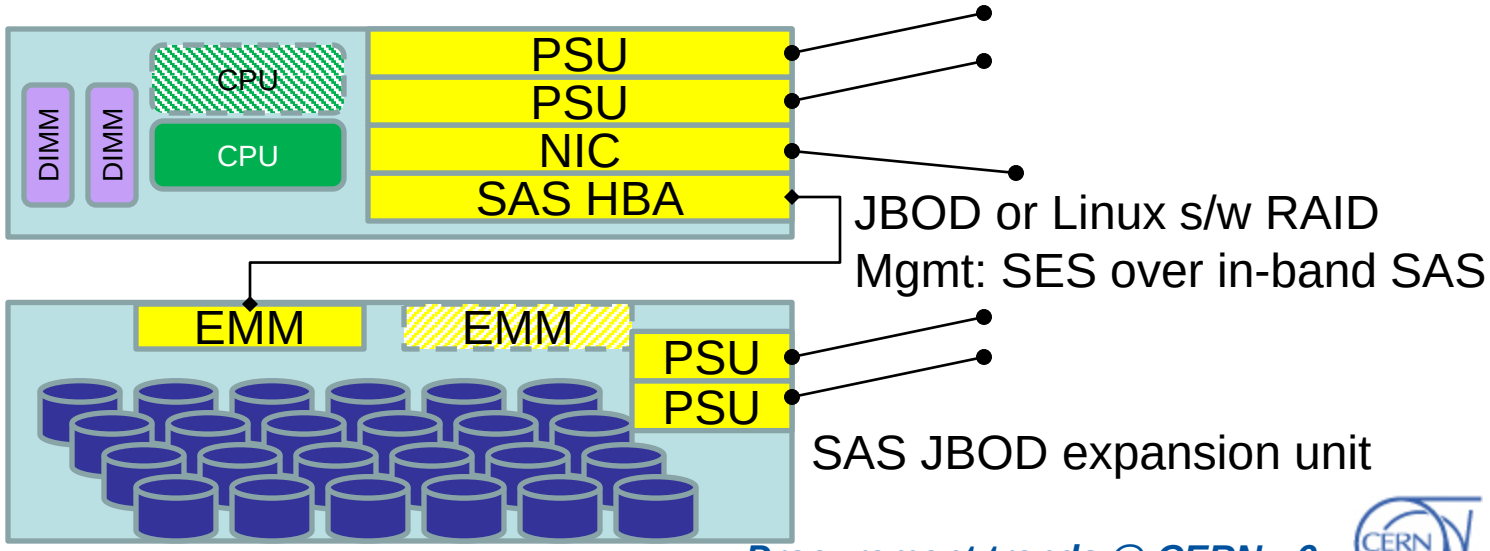


CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

New ‘disk server’

<u>Advantages</u>:
- Simplification: no hw RAID
- Flexibility: storage can be quickly re-connected to standby server
- Consolidation: server ≈ a normal batch server

<u>Disadvantages</u>:
- more power supply and cables
- SAS JBOD unit adds to price-tag



Solid State Drives

- Price/GB ~x10 high capacity SATA HDD
  - But difference is decreasing for higher perf HDD (~1:1 for 15krpm SAS)
- Even if affordable, replacing HDD with SSD isn't always a good idea
  - HDDs fails in an unpredictably way
  - SSDs 'wears out' in a predictably way strongly depending on data access type
- What does 'enterprise' really mean?
  - SLC only?
  - eMLC
  - MLC with special manufacturing attributes
    - Locked BOM (Built-Of Material)
    - Identical part can be tagged either consumer and enterprise!? check the warranty clause!

Solid State Drives @ CERN

CERN IT Department

- Tried to identify applicable use cases
- Batch worker nodes
  - Observed disk usage on batch workers is rarely exceeding 10% of available space…
  - Replaced HDD with SSD on a few batch workers.
    - Test result showed improved CPU usage 60% → 90% (peak)
  - Ordered a full batch 244 servers (dual E5-2630L), each with ~600GB SSD (MLC). Delivered last week…
- Tape buffer
  - Spindle contention in disk buffer degrades tape bandwidth (both read & write)
  - Put an SSD buffer in between?
    - Worst case scenario for a SSD with repeated streaming write-once-read-once access
  - Soon order two large (60bays) SAS JBODs fully populated with SSDs

On-site repairs

- Up to now: vendor provides the on-site warranty maintenance for delivered h/w
  - Usually subcontracted to local service companies
  - Service cost included in the bid adjudication
  - CERN procurement exclude ‘best value for money’ for supply contracts
- Some vendors tried to save money on warranty service
  - Delays: 30-50% missed repair targets
  - Mistakes: technicians lacks training
  - Overhead: servers not properly restored after repair
- We decided to move to off-site warranty
  - Sub-contract on-site maintenance repair server
  - Maintain stock and return broken parts for replacement

Hard disk headaches

- 2011 started with a huge campaign replacing 6792 2TB hard disks
  - In a Failure analysis on a dozen disks manufacturer admitted fabrication issues
  - Annoying long delay 5-6 month delay while servers were doing nothing
- Catastrophic flooding in Thailand
  - Apart for the terrible human and environmental consequences, hard disks market got a really hard hit
  - SSD became a serious option…

Foreseen changes in 2012 -

- Prepare for remote hosting
  - Tender for equipment we will never see or touch
  - Establish controlled workflows for
    - Remote hands unpacking & installation & cabling
    - Automatic network registration, initial boot up and burn-in
      - Deal with multiple NICs
      - Deal with wrongly connected NICs
      - Deal with wrong BIOS settings (boot list)
      - Deal with propagation of network registration credentials to a non-registered host
    - On-site hardware repairs, inventory and stock management

Conclusions

- In 2011 we reviewed our basic building blocks, architecture and processes
  - New disk server model will allow far-reaching consolidation of server procurement
  - New on-site warranty repair service
- In 2012 we anticipate most efforts to be concentrated on preparation for remote hosting in 2013.
  - Automation of machine registration, installation and testing
- In parallel, moving towards IaaS has become a CERN/IT priority
  - Decoupling the service dependency on bare hardware will greatly facilitates hw management

Questions?

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it