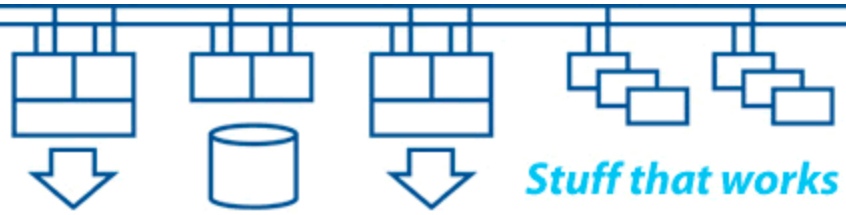
Itanium Solutions Alliance Innovation Awards

Mission-critical healthcare: NHSBT Pulse

Colin Butcher



What is Pulse?

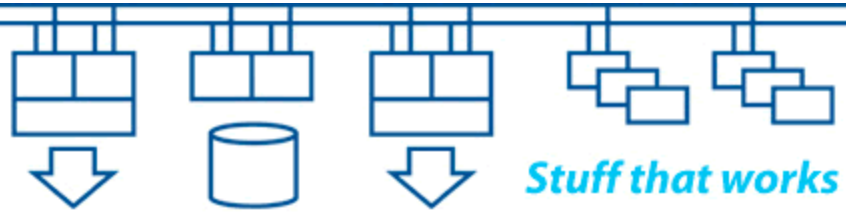
NHS Blood and Transplant

Pulse manages the collection, production and supply of blood products throughout England and North Wales

OCSL

Mimer

savant



Operational context

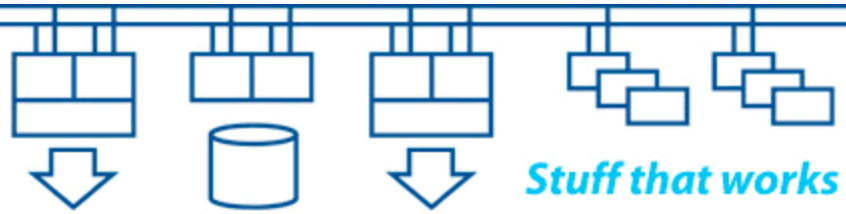
- Move from separate regional systems to single national system with consistent merged national database
- Increased availability and performance demands, hence need for disaster-tolerance
- Hardware refresh – systems, storage, infrastructure

  First had to determine if it was possible to migrate the data within an acceptable time window – so we built a proof of concept system to test it

Implementation specific context

- Pulse application updates (Savant)
- Database major version upgrade (Mimer)
- Operating system major version upgrade (OpenVMS)
- Platform migration from Alpha to Integrity (HP)
- Storage migration from HSG to EVA (HP)
- Data network infrastructure segmentation
- Common configuration and setup across all clusters
- Separate test & training environment

- All with minimal loss of service!



1) Design principles for mission-critical systems

An overview of the issues to be considered when designing, implementing and operating disaster-tolerant, mission-critical, multi-site systems.

DT, HA and DR – what do they mean?

- Disaster tolerance:
  - Continue operations while surviving major site outages without loss of data
- High availability:
  - Continue operations while surviving equipment and software failures without loss of data
- Disaster recovery:
  - The process of restarting sufficient operations to run the business after serious disruption

Risk

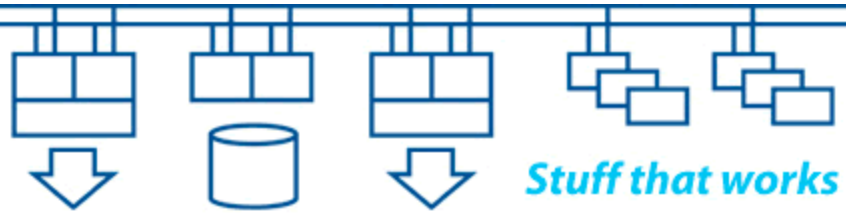
- What is the probability of a situation occurring?
- What is the impact if that situation occurs?
- What are the consequences?

- Most projects handle medium risk well enough
- Many projects over-specify to cater for what are in fact low probability or low impact issues
- Some projects under-specify and fail to cater for what are in fact high probability or high impact issues

Mission Critical systems

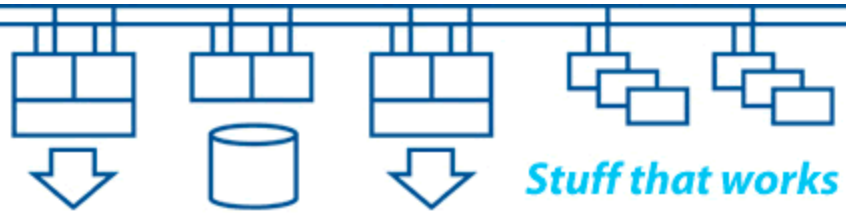
Mission critical systems need to be able to:

- Survive failures (resilience and failover)
- Survive changes (adapt and evolve)
- Survive people (simplify and automate)
- Never corrupt or lose critical data (data integrity)
- Requirements never remain static over an extended period of time, so we need to be able to make changes during the operational lifetime of the system
- Circumstances change, so we often need to be able to extend the operational lifetime and scope of a system
- It's not a theoretical exercise – it's real!



"Survivability test"

Cause of Outage:	Planned (Maintenance)	Unplanned (Failure)
Hardware	?	?
Operating System	?	?
Network Layer	?	?
Layered Products	?	?
Application Software	?	?
Application Data	?	?
Environment	?	?
People	?	?



RPO and RTO

How long have you got?

How much data can you afford to lose?

RTO = Recovery Time Objective

- What level of service outage can we tolerate?
- How quickly do we need to recover?
- How quickly do we need to be ready to deal with a subsequent failure?

RPO = Recovery Point Objective

- How much data can we tolerate losing?
- How quickly do we need to react to a failure?

Availability and performance

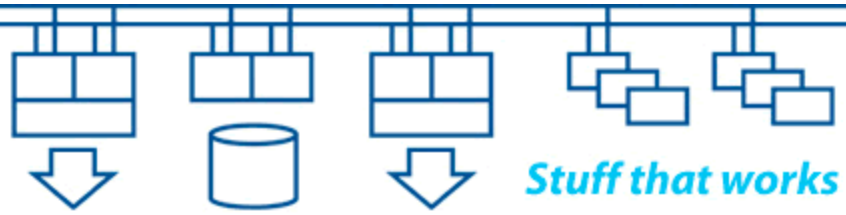
Availability:

- Probability of system being available for use at a given instant in time within the 'operational window'
- Function of both MTBF (reliability) and MTTR (repair time)

Performance:

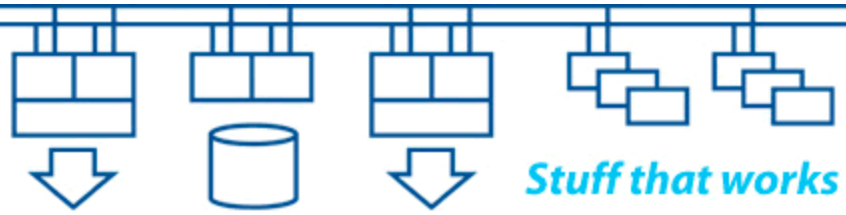
- Performance issues are often the cause of transient system failures and disruption
- The systems have to have sufficient capacity and performance to deal with the workload in an acceptable period of time under normal, failure and recovery conditions

Availability is more important than performance



Designing for availability

- Which parts of the system are mission-critical?
- Which parts of the system are safety-critical?
- What kind of failure do we prefer?
- What state transitions occur during failure and recovery?
- How can we recover from a failure without data loss or data corruption?
- How will you test your failure and recovery scenarios?

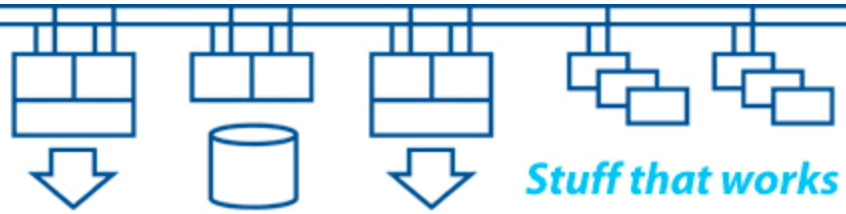


Understanding performance

- Bandwidth – determines throughput
  - It's not just "speed", it's throughput in terms of "units of stuff per second"
- Latency – determines response time
  - Determines how much "stuff" is in transit through the system at any given instant
  - "Stuff in transit" is the data at risk if there is a failure
- Jitter ("div latency" or variation of latency with time) – determines predictability of response
  - Understanding jitter is important for establishing timeout values
  - Latency fluctuations can cause system failures under peak load

Designing for performance

- Size systems to meet realistic criteria, eg: throughput; response time; minimal loss of "data in transit"
- Understand how the applications could break down into parallel streams of execution
- Understand scalability – do as much as possible once only, do little as possible every time
- Understand the need for synchronisation and serialisation of access to data structures
- Minimise "wait states" and contention

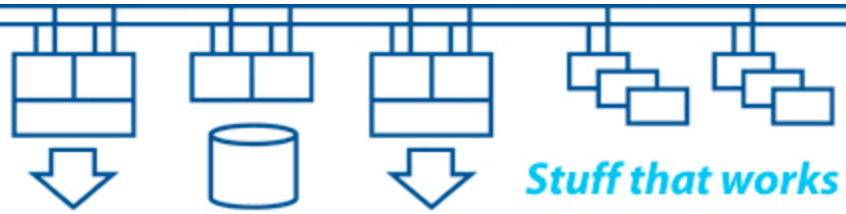


Multi-site issues

- Naming conventions
- Quorum and voting schemes
- Data replication schemes
- Effects of distance on network and storage protocols
- Symmetric or asymmetric operation – how good is your "crystal ball"?
- Centralised (and replicated) monitoring and alerting
- Remote access for management and operation
- Avoid automation of decision making

Testing

- We need to prove that service will continue with minimal disruption during failure and recovery
- We need to test for scale as well as functionality
- We need to test every aspect of the system and surrounding infrastructure under normal, failure and recovery conditions
- How will we generate a realistic load for testing?
- We need to regularly rehearse and test our procedures and plans to ensure that we stay current



2) NHSBT Pulse

An overview of how the new NHSBT Pulse systems fit into the NHSBT infrastructure.

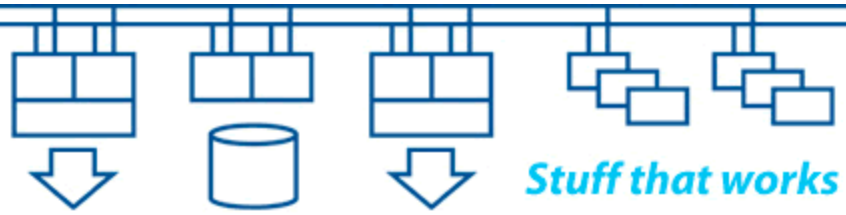
We have Production, Archive and Test environments with a shared common infrastructure, all of which is separated from the rest of the existing infrastructure.

Designed for future expansion and three-site working which allows us to move sites without downtime.

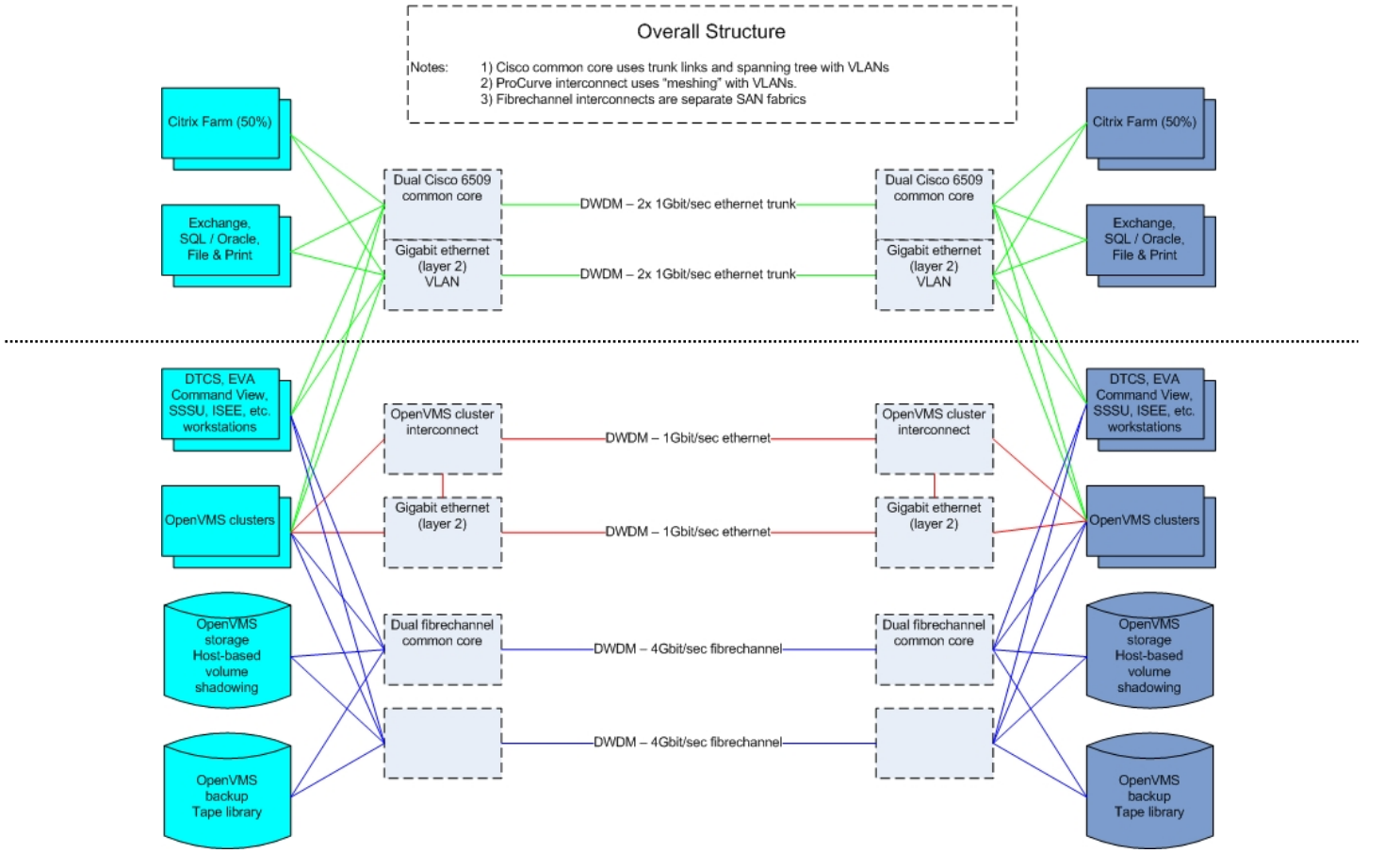
Functional layout

The systems are split up into:

- Common infrastructure (SAN fabrics, private network interconnects etc.)
- Production environment (a split-site cluster with host-based volume shadowed storage)
- Test environment (a functional replica of the Production environment)
- Archive environment (a single node cluster)
- Duplicated monitoring and reporting facilities
- External connectivity for users



Where these systems fit in the new infrastructure

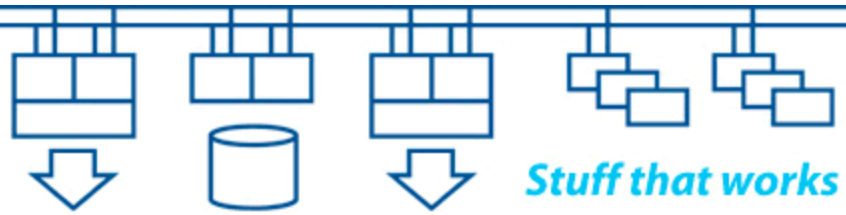


Hardware overview

- Integrity Servers (rx6600s and rx2660s)
- OpenVMS V8.3-1H1(plus patches)
- EVA 4100 storage arrays with 15k rpm 146GB drives
- MSL4048 tape libraries with Ultrium LTO4 FC drives
- SANswitch 4/32B fibrechannel switches with dual 2GigFC per fabric inter-site links
- ProCurve 3500yl-24 network switches with dual GigE inter-site links using ProCurve meshing
- Proliant DL380 G5 servers (DTCS monitoring, EVA command view, WEBES / ISEE reporting etc.)
- DTCS monitoring and alerting

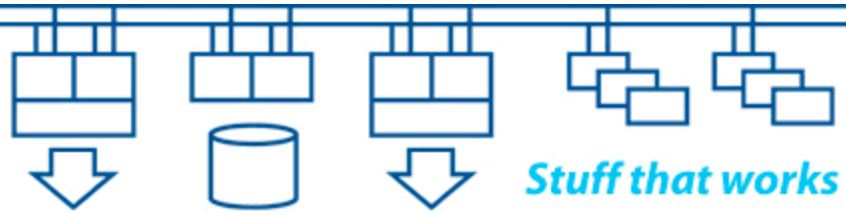
Integrity Servers (rx6600 and rx2660)

- Dual-core Itanium 2 CPUs
  - 3 socket / 6 core 1.6GHz 12MB cache in rx6600 (4 socket max.)
  - 1 socket / 2 core 1.6GHz 9MB cache in rx2660 (2 socket max.)
- 64GB in rx6600 (192GB max.)
- 16GB in rx2660 (32GB max.)
- 8 port built-in SAS array controller (2x IM arrays with hot spare)
- 4Gbps fibrechannel
  - 2x dual port HBAs in rx6600
  - 1x dual port HBA in rx2660
- 1Gbps ethernet
  - 4x fibre (user network), 4x copper (private interconnect) in rx6600
  - 2x fibre (user network), 4x copper (private interconnect) in rx2660
- iLO and serial console



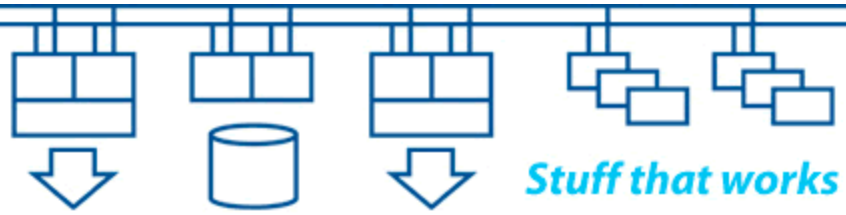
EVA 4100 storage arrays

- Single "disk group" using "double sparing"
- All spindles are FCAL 146GB 15k rpm
- All presented "Vdisks" are RAID 0 + 1
- SAN zoning ensures that
  - all EVAs are available to all Integrity Server systems
  - only Production EVAs are available to Production DL380s
  - only Test EVAs are available to Test DL380s
- Vdisk presentations control which Integrity Server systems can see which Vdisk devices within each EVA
- Mirrorclones and snapshots within the EVA are used to make copies for backups and other purposes
- Data replication using OpenVMS HBVS

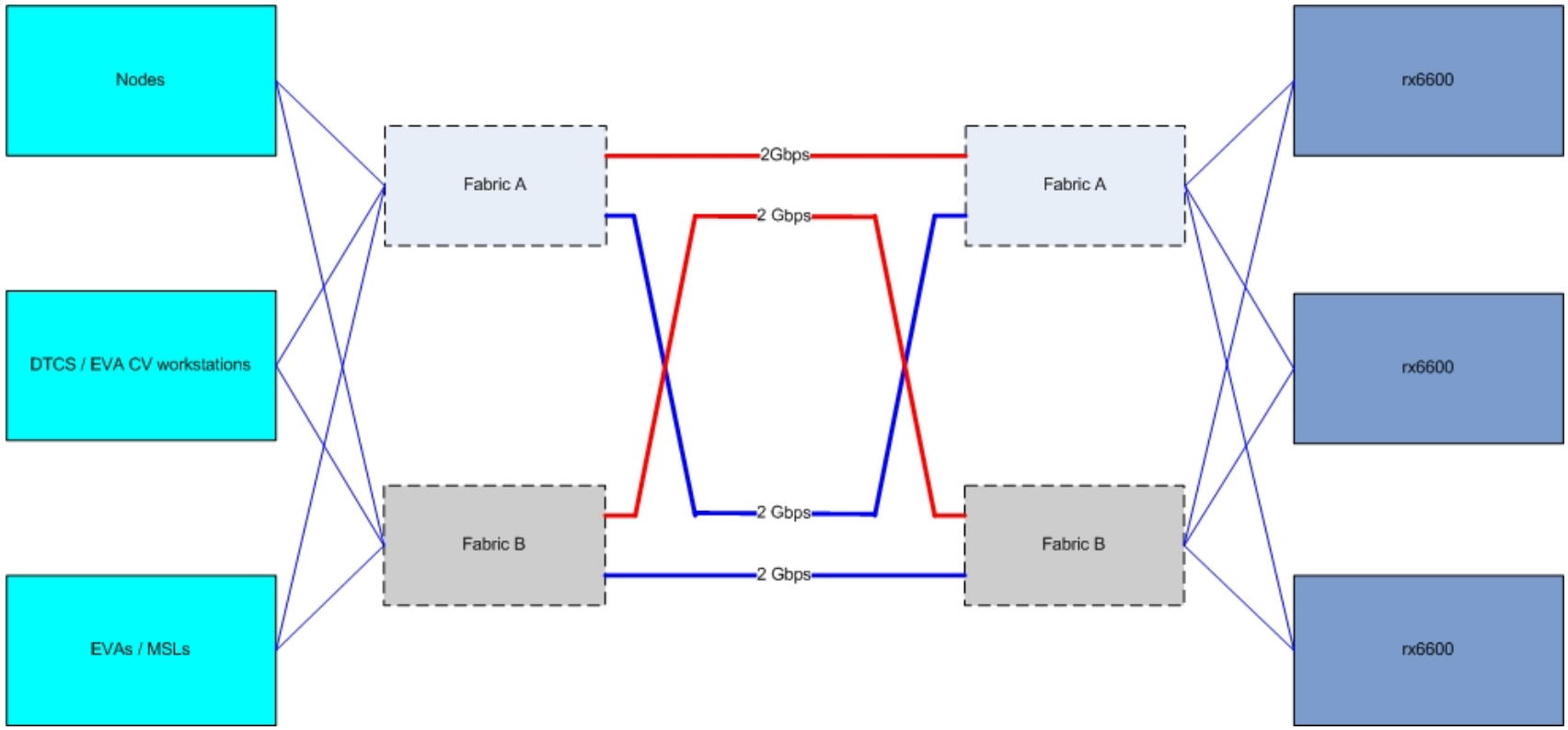


SAN switches and EVAs

- Dual-fabric SAN with no single point of failure
- All dual-port HBAs in all systems are dual path to two separate SANswitches
- SAN zoning and EVA presentations control access to devices
  - Only Production EVAs can be managed by Production EVA Command View workstations
  - Only Test EVAs can be managed by Test EVA Command View workstations
  - All initial EVA configurations were created using EVA scripts from the OpenVMS systems



Common SAN fabric infrastructure

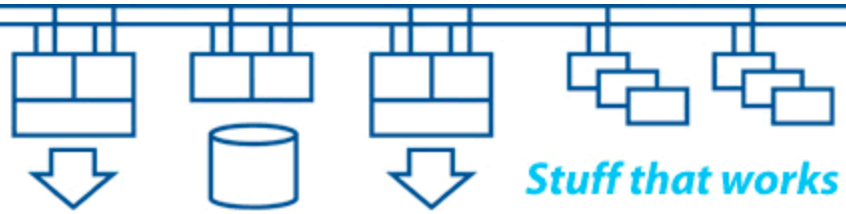


Network switches

- Ports grouped up for iLO / console devices, systems, cluster interconnects and inter-site links
- ProCurve meshing uses "shortest path" mechanism
  - Ports 23 & 24 use fibre for mesh links between switches
  - Ports 24 are the inter-site links using DWDM 1GigE links
- VLANs separate out the traffic types:
  - VLAN for TCP/IP to iLOs and SSSU scripting, AMDS for DTCS monitoring / quorum adjustment, DECnet for MDMS, LAT
  - VLAN for SCS path A
  - VLAN for SCS path B

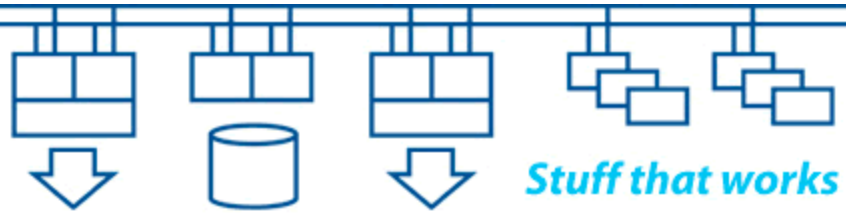
Protocols and VLANs

- Dual-rail VLANs over ProCurve mesh
  - SCS (locking, HBVS bitmap copying, etc.)
- LAN failover VLAN over ProCurve mesh
  - AMDS (DTCS monitoring and Quorum adjustment)
  - TCP/IP (SSSU access to EVA CV workstations, iLO access to all systems, iLO monitoring by DTCS, access to MSL4048s, access to SAN and network switches, firmware updates, etc.)
  - DECnet-Plus (MDMS, data copying etc.)
  - LAT (last-ditch terminal access)
- All NICs in all systems (except devices with a single port) are dual path to two separate ProCurve switches



External IP addressing

- All NICs in all systems are dual path to two separate Cisco switches (fibre, not copper)
- Uses "failsafe IP" for bandwidth and flexibility
- We use three kinds of IP address on the Cisco interfaces:
  - "Hidden" dedicated IP addresses for each NIC used for local reachability testing
  - Per-machine failsafe IP addresses used for systems management access
  - Application service failsafe IP addresses used for access to the applications and Databases. Disabled when not available. Only made available when the systems are ready for use.

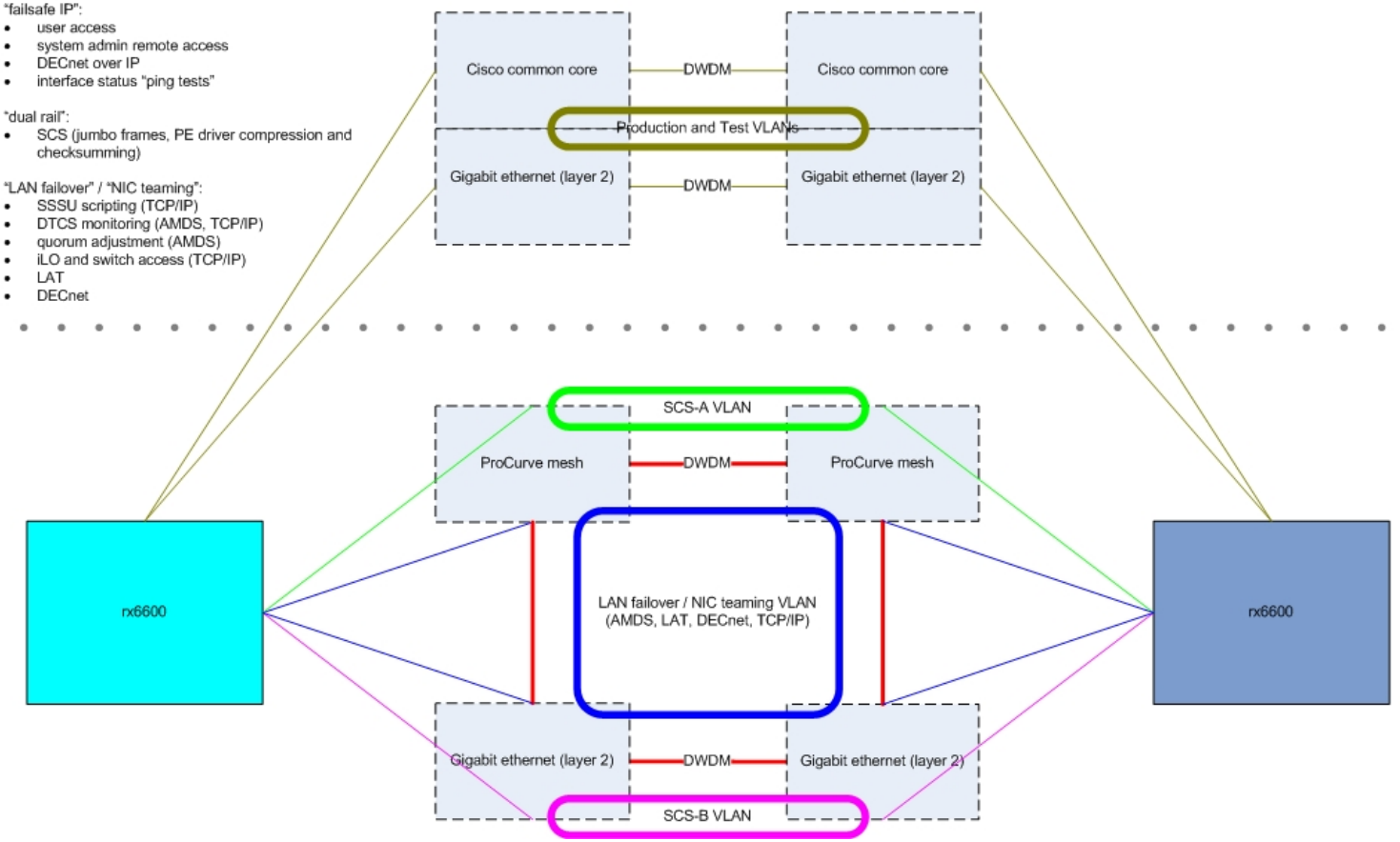


Common network infrastructure

"failsafe IP":
- user access
- system admin remote access
- DECnet over IP
- interface status "ping tests"

"dual rail":
- SCS (jumbo frames, PE driver compression and checksumming)

"LAN failover" / "NIC teaming":
- SSSU scripting (TCP/IP)
- DTCS monitoring (AMDS, TCP/IP)
- quorum adjustment (AMDS)
- iLO and switch access (TCP/IP)
- LAT
- DECnet



Multi-site clustering

- Split-site OpenVMS clusters give us "shared everything" access to data with protection from loss or corruption, even in the event of site failure
- Host-based volume shadowing (HBVS) ensures that data is consistent across all members of the shadow sets.
- The current quorum scheme lets Site A continue if Site B fails and protects us from data corruption due to a partitioned cluster
- The DTCS software monitors the systems for us and (most important of all) controls the formation of storage shadow sets when the systems boot and when nodes rejoin the cluster

Host-based volume shadowing (HBVS)

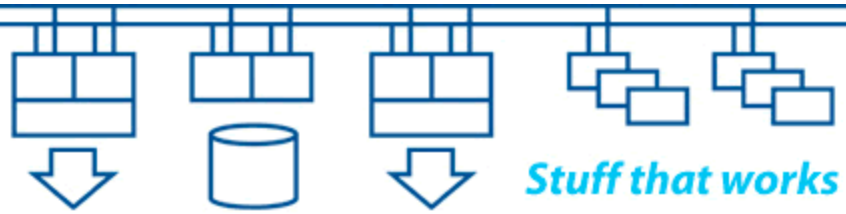
- The Production cluster uses 3 member shadow sets across 3x EVAs (2 EVAs at Site A, 1 EVA at Site B)
- The bootable system disk shadow sets at a site are only mounted by the nodes physically located at that site
- Local storage for page / swap / dump files
- The cluster-common disk is mounted by all nodes in the cluster and holds those files that must be unique and consistent across the entire cluster
- We make use of mini-copy and mini-merge by setting HBVS policies. These significantly speed up the catch-up process by maintaining write bitmaps
- Lots of small shadow sets give good granularity and control over HBVS behaviour

SAN boot for all cluster members

- The Integrity Server MP / EFI boot menus should be configured to disable auto-boot and to disable auto power-up following power loss
- The nodes boot from EVA presented Vdisk devices. Boot paths to SAN devices are configured using the BOOT_OPTIONS.COM mechanism by setting the boot device and the correct system root [SYSn.]
  Note: adding a new node to an existing cluster requires booting from another disk (eg: copy of the installation DVD) then mounting the target disc read-only
- Remove root [SYS0] to prevent anything booting accidentally into the cluster

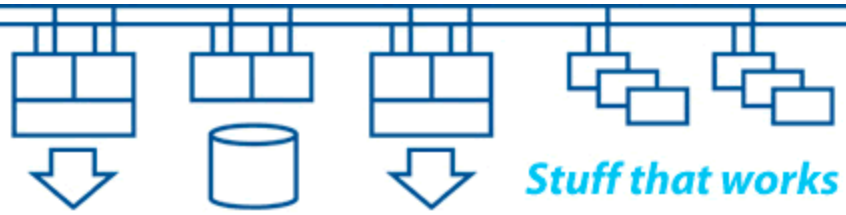
DTCS – monitoring and control

- DTCS is a set of HP and 3rd party products with installation, configuration and support services
- Remote console access, management and console output logging
- Integrated monitoring and quorum adjustment
- Rule based monitoring of individual systems / nodes
- Rule based SNMP polling of equipment
- Rule based TCP/IP "ping reachability" polling
- GUI and e-mail based alerting
- Scripting of failover and recovery actions across all systems / nodes and storage subsystems



3) Project delivery

An overview of why we believe the project was delivered successfully.



Success factors

- Small team of committed people
- Clear objectives and 'sensible' budget
- Built 'proof of concept' data migration system first
- Built system 'on paper', discussed it extensively and resolved potential technical problems prior to purchasing equipment and building system platform
- Project management and planning
- Leadership and collaborative working
- Trust between team members
- Sufficient flexibility to cope with issues as they arose

Thank you for your participation

Colin Butcher

For further information, please see: http://www.xdelta.co.uk/news#nhsbt