ATLAS T2/T3 WORKSHOP

Walker Stemple
November 2009

DELL CONFIDENTIAL

TOPICS

- Dell/Partner ATLAS Program Overview

Performance Optimization for HEP-SPEC

- BIOS optimization
- Subsystem evaluation and recommendations

Hardware

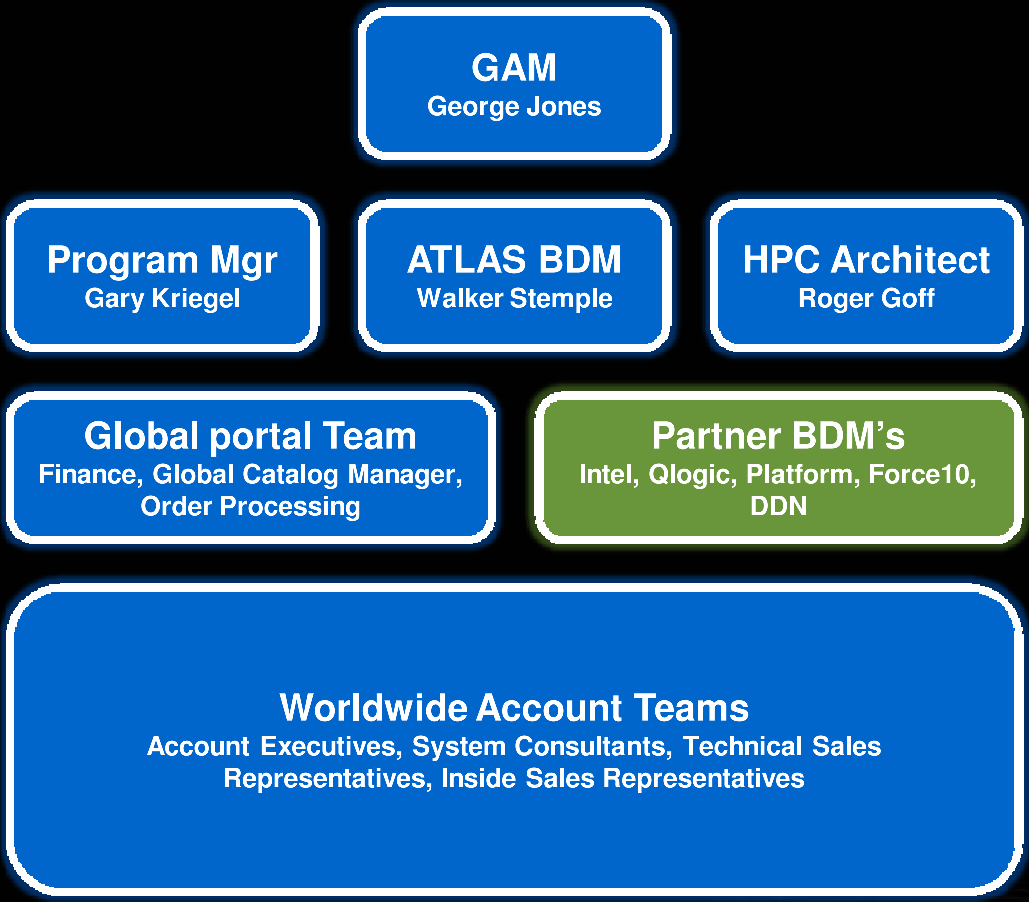
- Technical Review of Current Dell offerings:
  - Compute
  - Interconnect
  - Storage

Services

- Linux team
  - Scientific Linux
- Custom Fulfillment (aka “Merge Center”)

DELL ATLAS PROGRAM

Dell Scope



GAM
George Jones

Program Mgr
Gary Kriegel

ATLAS BDM
Walker Stemple

HPC Architect
Roger Goff

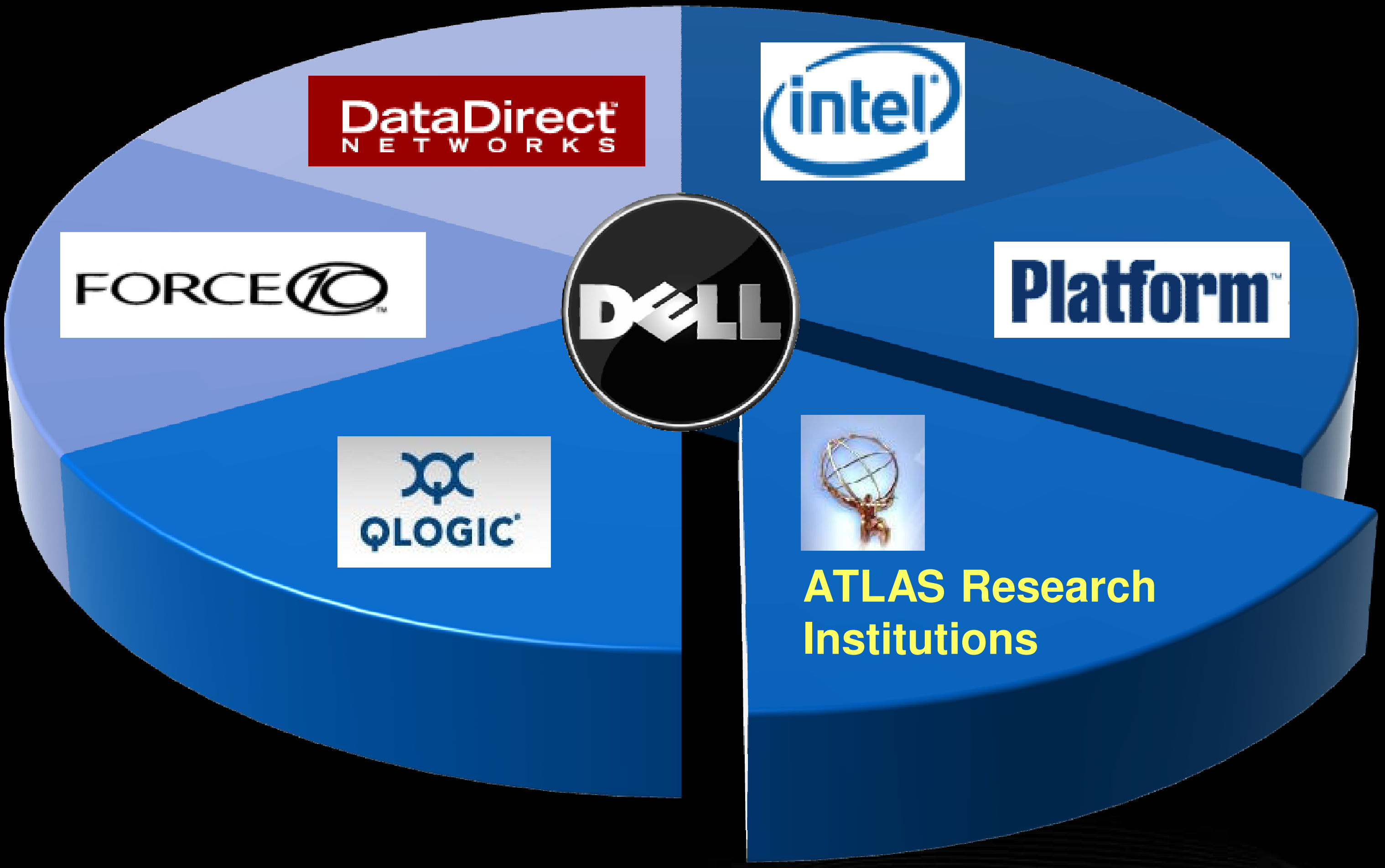
Global portal Team
Finance, Global Catalog Manager, Order Processing

Partner BDM's
Intel, Qlogic, Platform, Force10, DDN

Worldwide Account Teams
Account Executives, System Consultants, Technical Sales Representatives, Inside Sales Representatives

- Of the 299 Global ATLAS Institutions Dell has dedicated local account teams at 219
- 4820 ATLAS researchers worldwide supported by 13698 Dell Employees

DELL ATLAS COLLABORATION



ATLAS Research Institutions

RESEARCH COMPUTING AND THE DELL ATLAS PROGRAM

- Goal: To enable Science and Research at Universities and Gov't Institutions.

- The ATLAS Program is a 20 year project. Yes, Dell understands it is strange for a company to make a 20 year commitment as most companies struggle to determine what will happen in 90 or 180 days. But we are trying to think differently at Dell.

- Dell is providing Research Tools. Today the tools are Computer and Storage, and 20 years from now, we do not know what these tools will look like. But what we do know, is that worldwide , one Trillion dollars is spent on Research and Development each year. This number will continue to grow and there will be a need for Research Tools in 20 years. The Dell investment into the LHC/ATLAS project will help us strengthen our business and allow us to help further Science and Research.

ATLAS PERFORMANCE OPTIMIZATION

HEP-SPEC BENCHMARK

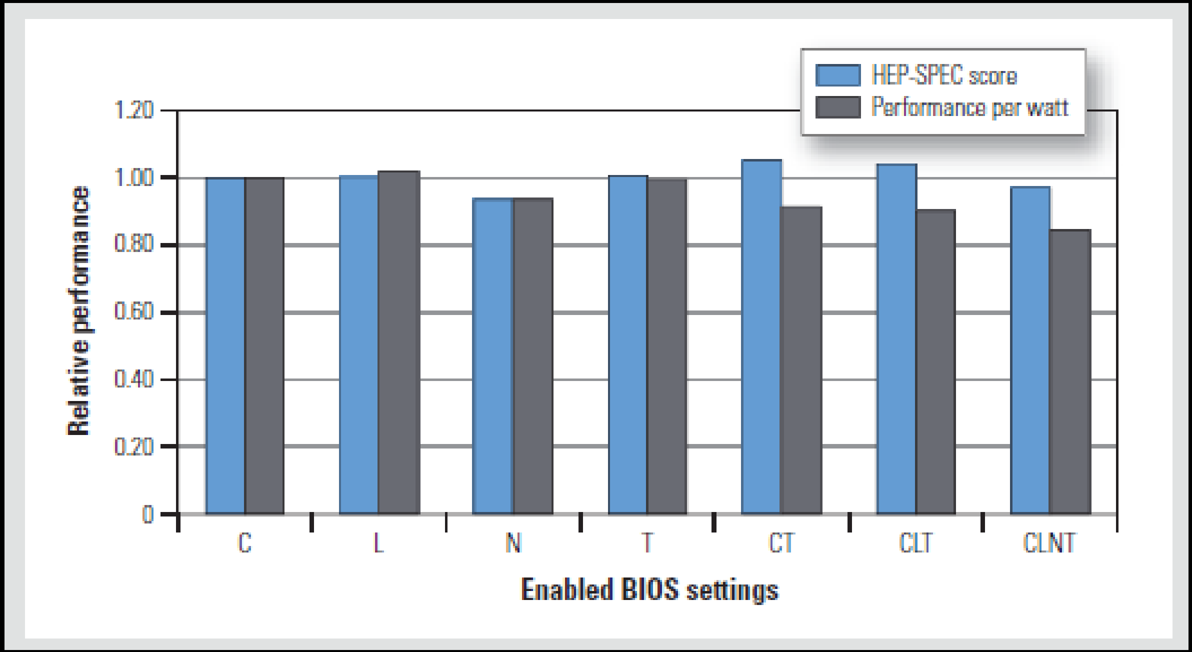
- Background
  - 2005 - Member institutions pledge compute resources in terms of benchmark units*
  - June 2007 - discovered that SPECint2000 does not scale linearly with High Energy Physics applications *
  - July 2008 - HEPIX Benchmarking WG proposed new benchmark based on the reasonable correlations seen with the four experiment applications *
  - May 2009 – Dell HPC Engineering teams begin to perform HEP-SPEC tests in support of Dell CERN/ATLAS Program

* Source:http://www.usatlasgrid.bnl.gov/twiki/bin/view/Admins/rsrc/Admins/CapacitySummary/New-CPU-Benchmark.pdf

HEP-SPEC OVERVIEW

- “High Energy Physics” benchmark
  - Based on SPEC CPU2006
    - spec_cpp, spec_rate subset
    - Static configuration (32 bit, GCC, -O2)
- Used for purchasing decisions
  - Results correlate to ATLAS online codes within 3-5%
  - Acquisitions based on SPEC units

11G BIOS OPTIONS



HEP-SPEC performance and efficiency relative to a system will all BIOS settings disabled

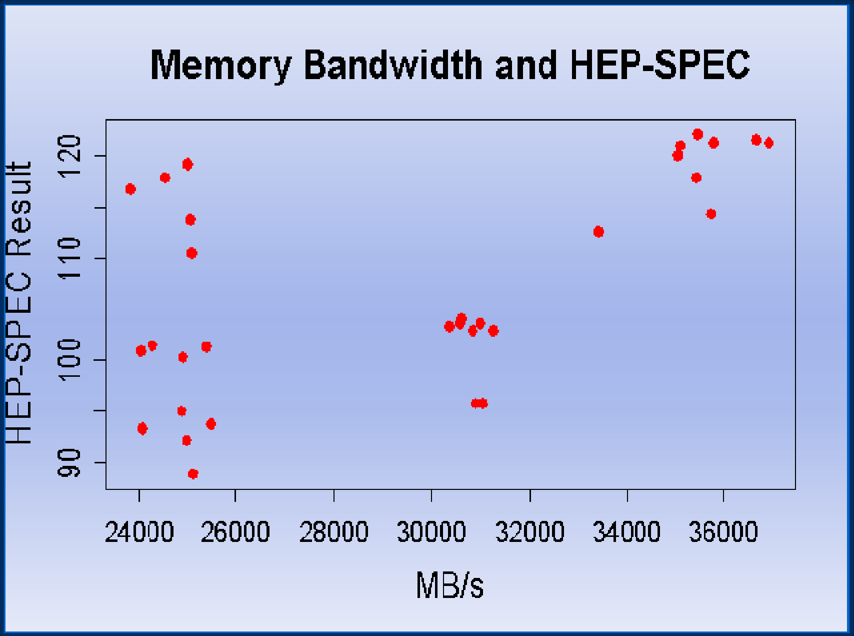
C-states: Allows the system BIOS to throttle power to individual processor cores based on need, which can enhance energy efficiency

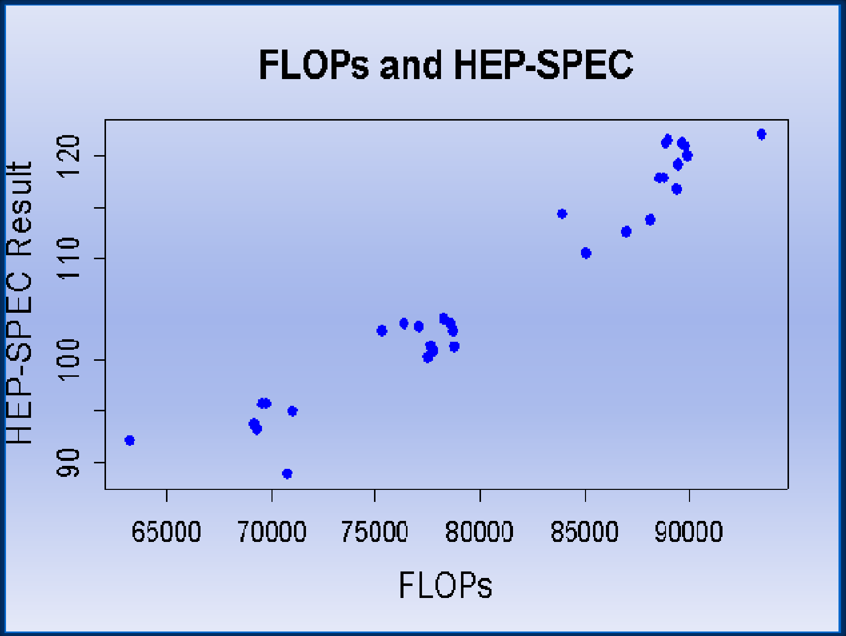
Logical processor (formerly called Intel Hyper-Threading Technology):
Improves thread-level parallelism by sharing the same physical core between multiple threads, which can increase performance for some codes

Node interleaving: Creates uniform memory access speed by interleaving memory across both processor sockets, which can help increase performance for codes that require a large global memory address space

Turbo mode: Increases processor clock rate by 1–3 increments of 133 MHz if there is available system power and heat headroom

SUBSYSTEM EVALUATION





- For dedicated HEP-SPEC computing resources, faster processors accelerate data processing more than faster memory.
- The HEP-SPEC performance difference between DIMM speeds is less than 3 percent

DELL RECOMMENDED PROCESSOR/MEMORY PAIRINGS FOR HEP-SPEC PERFORMANCE

Application Priority	Processor	Memory Configuration
Energy efficiency	L5520	4x4GB @ 1066MHz
Absolute Performance	X5570	6x4GB @ 1333MHz
Balanced	E5540	6x4GB @ 1066MHz
Value	E5520	4x4GB @ 1066MHz
Mixed Workloads	X5550	6x4GB @ 1066MHz

FUTURE RESEARCH

- Evaluate ATLAS production codes:
  - Online
    - Level 2 Trigger
    - Event Filter
  - Offline
    - Simulation
    - Reconstruction
    - Analysis

QUESTIONS?

DELL TECHNOLOGY FOR CERN/ATLAS

- Compute
- Interconnect
- Storage

COMPUTE

POWEREDGE SERVER: PORTFOLIO

R410: 2S

1u HPC Rack

R610: 2S

1u Rack

R710: 2S

2u Rack

M610: 2S Half Height Blade

M710: 2S Full Height Blade

- Unprecedented RELIABILITY and COMMONALITY
- Distinguished, PURPOSEFUL Design
- Industry leading system EFFICIENCY

Rack Compute Node - POWEREDGE R410

		PREVIOUS		LATEST
PERFORMANCE		PE 1950 III	PE SC1435	PE R410
	CHIPSET	GreenCreek	Broadcom	Intel
	PROCESSOR	Harpertown, Wolfdale	AMD	Intel
AVAILABILITY	SOCKET	2S	2S	2S
	MEMORY	8 x FBD	8 X DDR2	4+4 DDR3
	DIMM CAPACITY	512MB, 1, 2, 4, 8 GB	512MB, 1, 2, 4 GB	1, 2, 4, 8 GB
	SLOTS	2 PCIe x 8 or PCI-x	1x PCIe x8 and 1x PCI-X	1x PCIe x16
EXPANDABILITY	HDD	2 x 3.5” or 4 x 2.5”	2 x 3.5”	4x 3.5” (optional 2.5”)
	HDD	HotPlug	Cabled	Optional Hot Swap
	POWER SUPPLY	HotPlug, Redundant	Non-RDNT	Optional RDNT
	LOM	2 x TOE	2 GbE	2 GbE
	DIAGNOSTIC	LCD	Quadpack LED	Quadpack LED, optional LCD
	MANAGEMENT	BMC+DRAC 5	BMC	BMC , IPMI 2.0 Compliant Optional iDRAC6-Express and iDRAC6-Enterprise
	Internal STORAGE	Yes, Unmanaged	NO	2 x Internal USB
	SECURITY	TPM 1.2	NO	TPM
	POWER SUPPLY		600W	480W / 500W
	CLIMATE SAVER			YES

PowerEdge R610

		CURRENT	FUTURE
PERFORMANCE		PowerEdge 1950 III	PowerEdge R610
	CHIPSET	Greencreek	Tylersburg
	PROCESSOR	Harpertown, Wolfdale	Nehalem
AVAILABILITY	SOCKET	2S	2S
	MEMORY SLOTS	8 x FBD	12 x DDR3
	DIMM SIZES	512 MB, 1, 2, 4, 8 GB	1, 2, 4, 8 GB
	EXPANSION SLOTS	2 PCIe x 8 or PCI-x	2 PCIe Gen 2
	LOM	2 x TOE	4 x TOE
EXPANDABILITY	HDD	2 x 3.5” or 4 x 2.5”	6 x 2.5”
	HDD	Hot Plug	Hot Plug
	POWER SUPPLY	Hot Plug, Redundant	Hot Plug, Redundant
MANAGEMENT	COOLING	Redundant	Redundant
	DIAGNOSTIC	LCD	LCD
	MANAGEMENT	BMC+DRAC 5	Advanced Manageability
	PERSISTENT STORAGE	Yes, Unmanaged	Yes, Managed
	SECURITY	TPM 1.2	TPM 1.2

POWEREDGE R710

		CURRENT	FUTURE
PERFORMANCE		PowerEdge 2950 III	PowerEdge R710
	CHIPSET	Greencreek	Tylersburg
	PROCESSOR	Harpertown, Wolfdale	Nehalem
AVAILABILITY	SOCKET	2S	2S
	MEMORY SLOTS	8 x FBD	Up to 18 x DDR3
	DIMM SIZES	512 MB, 1, 2, 4 GB	1, 2, 4, 8 GB
	EXPANSION SLOTS	3 PCIe or 2 PCI-x + 1PCIe	2 PCIe x8 + 2 PCIe x4 G2 Or 1 x16 + 2 x4 G2
	LOM	2 x TOE	4 x TOE
EXPANDABILITY	HDD	6 x 3.5” or 8 x 2.5”	6 x 3.5” or 8 x 2.5”
	HDD	Hot Plug	Hot Plug
	POWER SUPPLY	Hot Plug, Redundant	Hot Plug, Redundant
MANAGEMENT	COOLING	Hot Plug, Redundant	Hot Plug, Redundant
	DIAGNOSTIC	LCD	LCD
	MANAGEMENT	BMC+DRAC 5	Advanced Manageability
	PERSISTENT STORAGE	Yes, Unmanaged	Yes, Managed
	SECURITY	TPM 1.2	TPM 1.2

INTERCONNECT

POWERCONNECT 6000 SERIES

MANAGED ROUTING GIGABIT SWITCHES W / 10GE

PowerConnect

6248 & M6220

- **Flexibility**
  - •24 and 48 port Gigabit with PoE or Fiber dense options, all with 4 10GE modular ports

Performance & Reliability

- Wire speed across all ports
- Redundant Power Optional

Routing

- RIP, OSPF, VRRP, IP Multicast

Security

- Access Control Lists, MS NAP, 802.1x, Auto VLAN

- **Stacking**
  - Unified management
  - Up to 12 switches or 576 ports
  - 48Gb redundant architecture

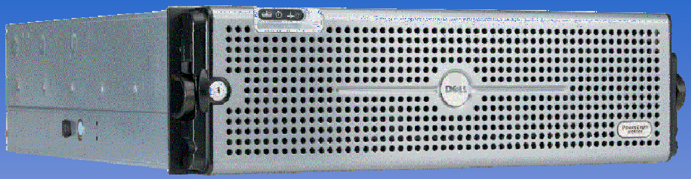
Four Modular 10 GE Ports

- Available Modules:
  - 10GbaseT (Q1-08)
  - SFP+ (Q3-08)
  - Resilient Stacking (48Gb)
  - XFP
  - CX4

STORAGE

POWERVAULT MD1000

Simple Server Expansion with Compelling Cost per GB



Key Attributes

- **Size** - *3U JBOD with 15 drives per shelf*
- **Drive Flexibility** – *SAS and SATA in a single enclosure*
- **Capacity** – *Up to 90 drives when attached to PERC 6E*

Simple Storage Expansion

- Expansion for PowerEdge server and PowerVault MD3000 and MD3000i
- Expands to 90 drives behind a PERC 6/E RAID controller.

Drive Flexibility

- Support for both SAS and SATA disk drives in a single enclosure.
  - SAS, Nearline SAS, SATA and Energy Efficient SATA

PERC RAID Controller

- PERC 6/E RAID controller provides enhanced performance, ease of use and reliability over previous generations.

Optimized for PowerEdge Server Environments

- Manage internal and external storage via a common interface.

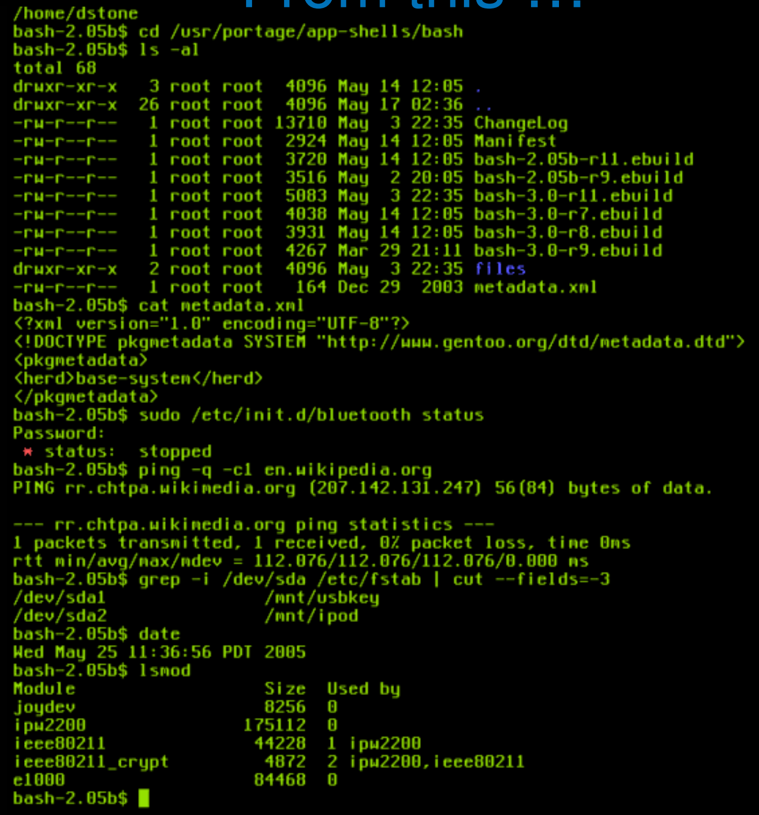
SOFTWARE

Cluster Management – Dell Edition

Platform™

Alternative open source solution

From this …

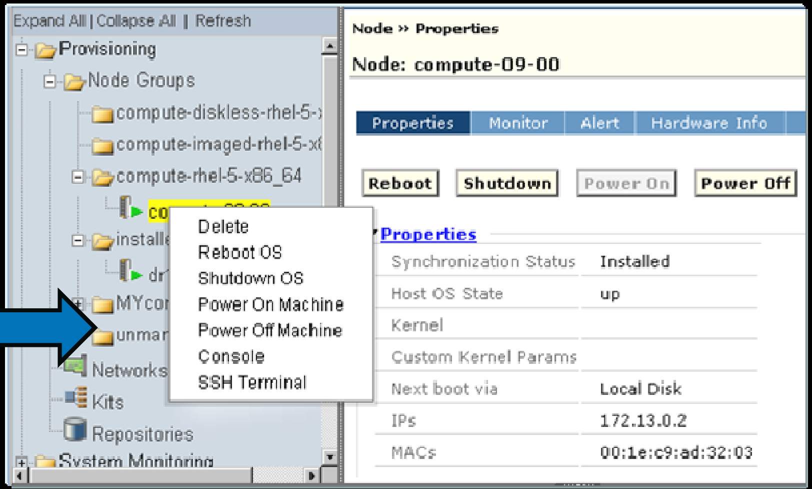


```
/home/dstone
bash-2.05b$ cd /usr/portage/app-shells/bash
bash-2.05b$ ls -al
total 68
drwxr-xr-x   3 root root  4096 May 14 12:05 .
drwxr-xr-x  26 root root  4096 May 17 02:36 ..
-rw-r--r--   1 root root 13710 May  3 22:35 ChangeLog
-rw-r--r--   1 root root  2924 May 14 12:05 Manifest
-rw-r--r--   1 root root  3720 May 14 12:05 bash-2.05b-r11.ebuild
-rw-r--r--   1 root root  3516 May  2 20:05 bash-2.05b-r9.ebuild
-rw-r--r--   1 root root  5083 May  3 22:35 bash-3.0-r11.ebuild
-rw-r--r--   1 root root  4038 May 14 12:05 bash-3.0-r7.ebuild
-rw-r--r--   1 root root  3931 May 14 12:05 bash-3.0-r8.ebuild
-rw-r--r--   1 root root  4267 Mar 29 21:11 bash-3.0-r9.ebuild
drwxr-xr-x   2 root root  4096 May  3 22:35 files
-rw-r--r--   1 root root   164 Dec 29  2003 metadata.xml
bash-2.05b$ cat metadata.xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE pkgmetadata SYSTEM "http://www.gentoo.org/dtd/metadata.dtd">
<pkgmetadata>
<herd>base-system</herd>
</pkgmetadata>
bash-2.05b$ sudo /etc/init.d/bluetooth status
Password:
 * status:  stopped
bash-2.05b$ ping -q -c1 en.wikipedia.org
PING rr.chtpa.wikimedia.org (207.142.131.247) 56(84) bytes of data.

--- rr.chtpa.wikimedia.org ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 112.076/112.076/112.076/0.000 ms
bash-2.05b$ grep -i /dev/sda /etc/fstab | cut --fields=-3
/dev/sda1               /mnt/usbkey
/dev/sda2               /mnt/ipod
bash-2.05b$ date
Wed May 25 11:36:56 PDT 2005
bash-2.05b$ lsmod
Module                  Size  Used by
joydev                  8256  0
ipw2200               175112  0
ieee80211              44228  1 ipw2200
ieee80211_crypt         4872  2 ipw2200,ieee80211
e1000                  84468  0
bash-2.05b$ █
```

PCM – Dell Edition

To this …

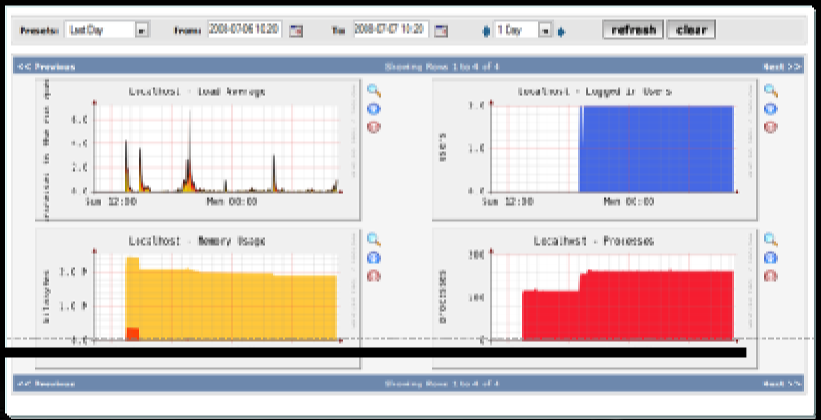


Node » Properties

Node: compute-09-00

Properties

Synchronization Status	Installed
Host OS State	up
Kernel	
Custom Kernel Params	
Next boot via	Local Disk
IPs	172.13.0.2
MACs	00:1e:c9:ad:32:03



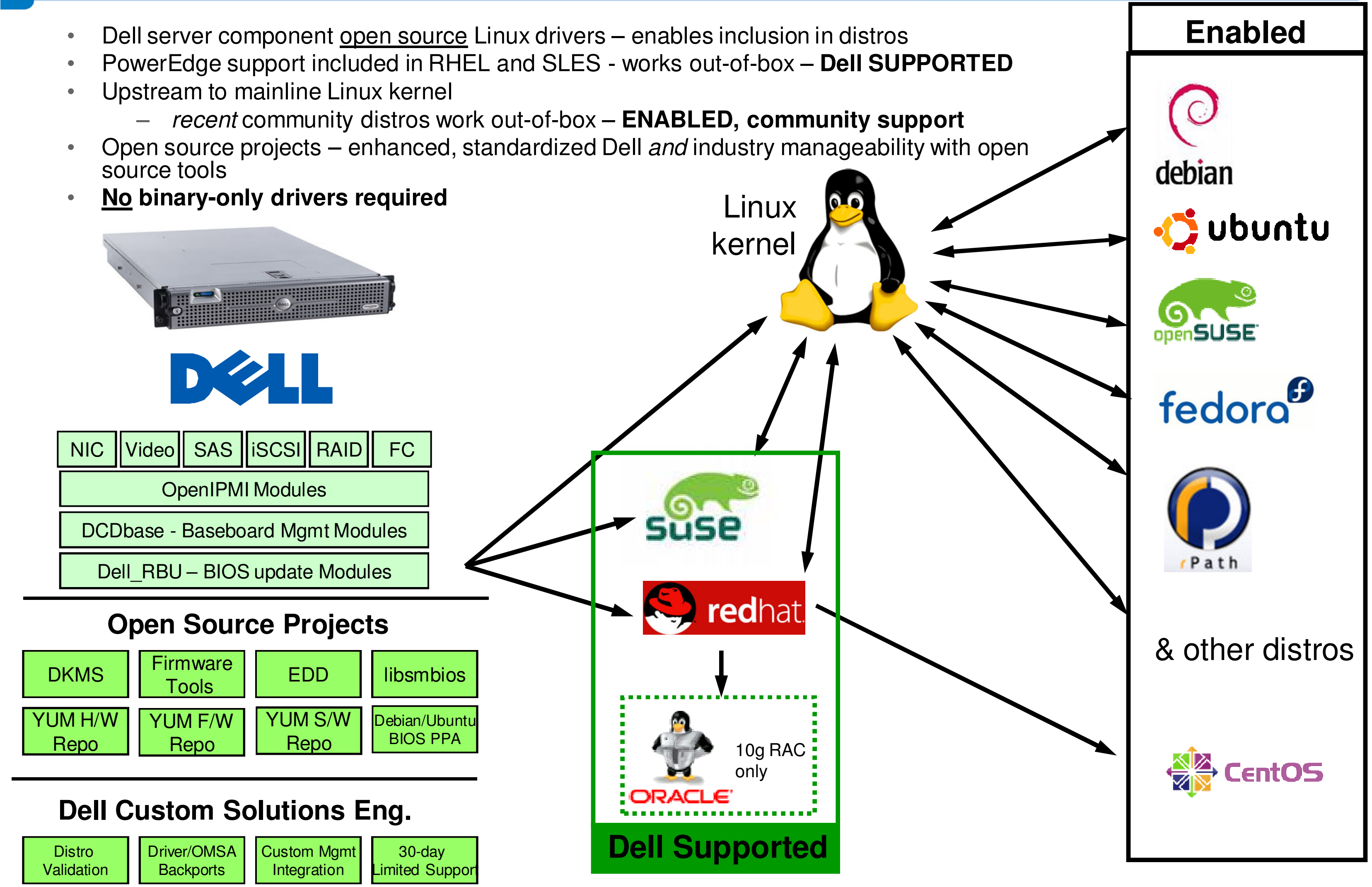
The PCM – Dell Edition provides a web-based interface that makes HPC clusters easy to manage

SERVICES

LINUX ENGINEERING

Dell Linux Development Strategy - Open Source Leverage

- Dell server component open source Linux drivers – enables inclusion in distros
- PowerEdge support included in RHEL and SLES - works out-of-box – **Dell SUPPORTED**
- Upstream to mainline Linux kernel
  - *recent* community distros work out-of-box – **ENABLED, community support**
- Open source projects – enhanced, standardized Dell *and* industry manageability with open source tools
- **No binary-only drivers required**



Enabled

debian

ubuntu

openSUSE

fedora

rPath

& other distros

CentOS

Linux kernel

Dell Supported

SuSE

redhat

10g RAC only

ORACLE

NIC	Video	SAS	iSCSI	RAID	FC

OpenIPMI Modules

DCDbase - Baseboard Mgmt Modules

Dell_RBU – BIOS update Modules

Open Source Projects

DKMS	Firmware Tools	EDD	libsmbios
YUM H/W Repo	YUM F/W Repo	YUM S/W Repo	Debian/Ubuntu BIOS PPA

Dell Custom Solutions Eng.

Distro Validation	Driver/OMSA Backports	Custom Mgmt Integration	30-day Limited Support

RHEL / SL 5

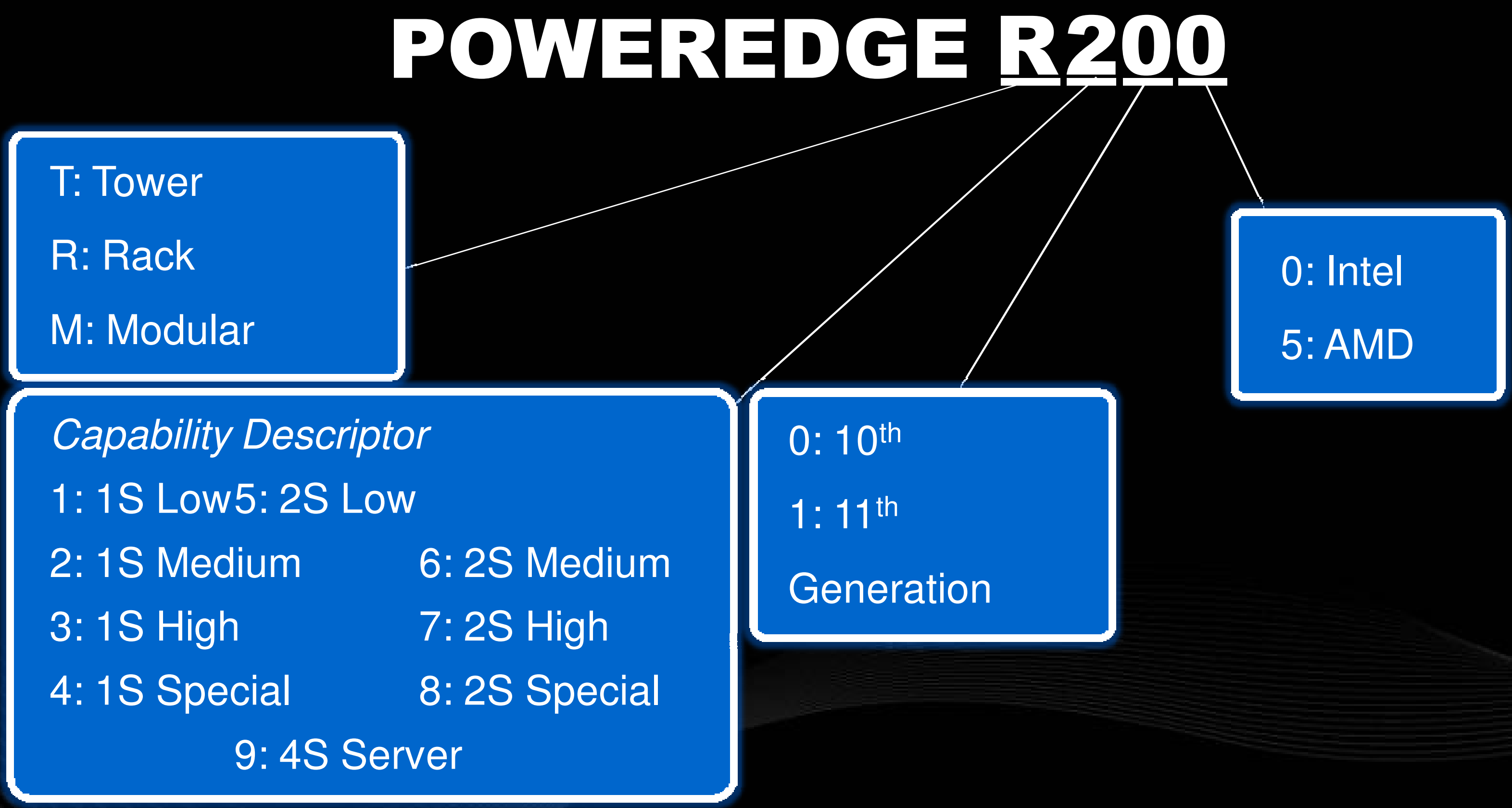
- Scientific Linux (SL) is a RHEL clone, i.e. it’s simply recompiled open source RHEL source code with the RH trademarks stripped out, and some added applications (eg. FITS libraries, Graphviz, and R.). Since Dell open sources all our device drivers and OpenManage systems management OS kernel modules, which we directly integrate into RHEL, therefore all the device driver compatibility, validation, OMSA agents and other eng. efforts we put into RHEL are inherited by SL (and CentOS), i.e. just as RHEL “just works” on Dell all servers, so does SL.

- We provide informal support for community distros like SL, Debian, Fedora, etc. through community mailing lists at http://linux.dell.com

- Lastly, SL administrator can install OMSA agents from the Dell YUM repository at http://linux.dell.com/repo/hardware . The YUM repository also contains BIOS/firmware updates packaged as .RPMs, as well as device driver updates. This means SL admins can use native Linux yum and SL OS update commands to also do Dell PowerEdge server (and client) hardware patch management.

POWEREDGE SERVER NAMING

- Provides better understanding of Dell’s server portfolio
- Allows for quick comparisons between Tower, Rack and Modular server capabilities



POWEREDGE R200

T: Tower

R: Rack

M: Modular

0: Intel

5: AMD

Capability Descriptor

1: 1S Low5: 2S Low

2: 1S Medium 6: 2S Medium

3: 1S High 7: 2S High

4: 1S Special 8: 2S Special

9: 4S Server

0: 10th

1: 11th

Generation

POWEREDGE R810

SCALABLE 2/4 -SOCKET 2U SERVER FOR PERFORMANCE DENSITY

Overview

- High performance, high reliability, flexible 2U server that scales to 4 Sockets.
- Best for use as a email messaging, medium-database, or virtualization server.
- High CPU Core Count and Memory Capacity.

Benefits

- Cost effective scaling and better price per performance than mainstream 2S/4S servers.
- Easy manageability with enterprise class system management tools including Lifecycle Controller via iDRAC Express or Enterprise upgrade
- Maximize datacenter density and performance.

Performance

- Up to Eight-Core Intel Nehalem EX processors
- 32 DDR3 DIMM slots for a total of 512GB of RAM
- PCI-Express I/O Technology

Availability

- Hot-plug SAS or SATA hard drives
- Memory: ECC
- Hot-plug, redundant power and cooling
- Baseboard Management Controller with IPMI 2.0
- Optional remote management (iDRAC6)

Expandability, I/O, Storage

- 6 PCI slots PCI-E Gen 2
- Optional PERC7i/SAS7iR
- Configuration options with 6 HDD

Simplified Systems Management

- Baseboard Management Controller with IPMI 2.0
- Advanced management functionality with Lifecycle Controller enabled via optional upgrade to iDRAC Express or Enterprise
- Interactive LCD for easy monitoring and diagnostics

Rack File Server - POWEREDGE R510

		CURRENT		FUTURE
PERFORMANCE		PowerEdge 2950 III	PowerEdge R710	PowerEdge R510
	CHIPSET	Greencreek	Tylersburg 36	Tylersburg 24
	PROCESSOR	Harpertown/ Wolfdale	Nehalem	Nehalem
AVAILABILITY	SOCKET	2S	2S	2S
	MEMORY	8 x FBD	Up to 18 x DDR3	Up to 8 x DDR3
	DIMM CAPACITY	512 MB, 1, 2, 4 GB	1, 2, 4, 8 GB	1, 2, 4, 8 GB
	SLOTS	3 PCIe or 2 PCI-x + 1PCIe	2 PCIe x8 + 2 PCIe x4 G2 Or 1 x16 + 2 x4 G2	3 PCIe x8 + 1 Internal Storage Slot Or 1 x16 + 1 Internal Storage Slot
EXPANDABILITY	HDD	6 x 3.5" or 8 x 2.5"	6 x 3.5" or 8 x 2.5"	4 or 8 or 12+2 3.5" or 2.5"
	HDD	Hot-swap	Hot-swap	Cabled or Hot-swap
	POWER SUPPLY	Hot-swap	Hot-swap, RDNT	Hot-swap, RDNT
	LOM	2 x TOE	4 x TOE	2 GbE
	DIAGNOSTIC	LCD	LCD	LED/LCD
	MANAGEMENT	BMC+DRAC 5	Advanced Manageability	BMC+ Opt. iDRAC6 Express/ Enterprise
	PERSISTENT STORAGE	2x Internal USB	Yes, Managed	2 x Internal USB
	SECURITY	TPM	TPM	TPM

TOWER COMPUTE NODE POWEREDGE T410

Replacement for
PowerEdge 1900 & T605

PERFORMANCE:

- Up to two Intel Nehalem processors
- 4+4 DDR3 (with 8GB DIMMs)

AVAILABILITY:

- Six 2.5” or 3.5” hot-plug optional hard drives
- Management: iDRAC (Express/Enterprise)
- TPM 1.2, IPMI, BMC

EXPANDIBILITY, I/O & STORAGE

- 5 PCIe Slots (1x16x8 and 4x8x4)
- 2 Embedded Gigabit NICs with TOE
- 2 Internal USB for Persistent Storage

POWERCONNECT 8000 SERIES

MANAGED ROUTING 10GE SWITCHES

Fall '09

8024 & 8024F & M8024

PowerConnect

High Density

- 24 ports of 10 Gigabit in 1U
- Modular Switch – M8024
  - 16 internal ports for blade servers
  - Up to 8 external ports in two modules

Routing - RIP, OSPF, VRRP, IP Multicast

8024 and 8024F

Dense Flexibility

- 8024 – 24 10GBASE-T ports with 4 SFP+ Combo
- 8024F – 24 SFP+ ports with 4 10GBASE-T Combo

High Availability

- Hot Swap Power/Fans
- Dual FW Images

M8024

Flexible Connectivity

- 4 port SFP+ Module
- 3 port CX4 Module
- 2 port 10GBASE-T

Highly Manageable

- Simple Switch mode for fast, flexible deployment
- Managed via CMC, SNMP, CLI, GUI

Simple Switch

Enables server administrators to deploy high performance network switching without having to engage a network admin to set up the blade network, saving time and money

CUSTOM FUFILLMENT

DELL CUSTOM FULFILLMENT SERVICES

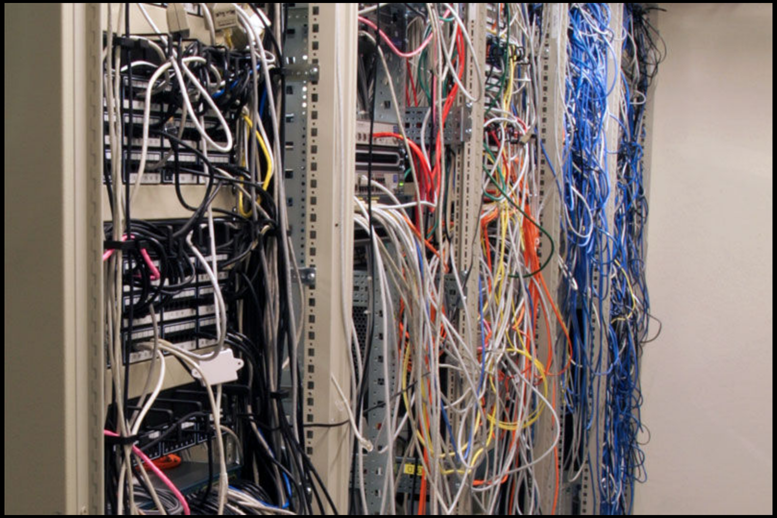
STATE-OF-THE-ART PROCESS

- 130,000 sq ft of operating space
- ISO 9001-certified process
- Onsite project management, scalable direct workforce

A HISTORY OF EXCELLENCE

- Started in the late 90's with one customer and single offering
- 425 customers and 30+ standard offerings in FY08
- Offerings have evolved in direct response to customer requirements
- FY08 Output:
  - 1.5 million boxes
  - 750,000 systems

HARDWARE AND RACK & STACK



CONVENTIONAL HARDWARE DEPLOYMENT

- Onsite IT receives hardware, server components and parts
- Multiple-day system and server rack construction and configuration
- Inconsistent construction, configuration and quality

HARDWARE INTEGRATION & CONFIGURATION

- Parts installed/de-installed on servers, systems and storage
- Power-on testing and configuration
- Available on Dell and non-Dell systems

RACK AND STACK CONSTRUCTION

- Dell-standard quality integration, cabling and labeling
- Consistency, quality and white-glove delivery

THANK YOU!

Walker Stemple
Dell Global CERN/ATLAS Business Development Manager
+1.512.239.9537 | Walker_Stemple@dell.com
www.dell.com/ATLAS

BACKUP