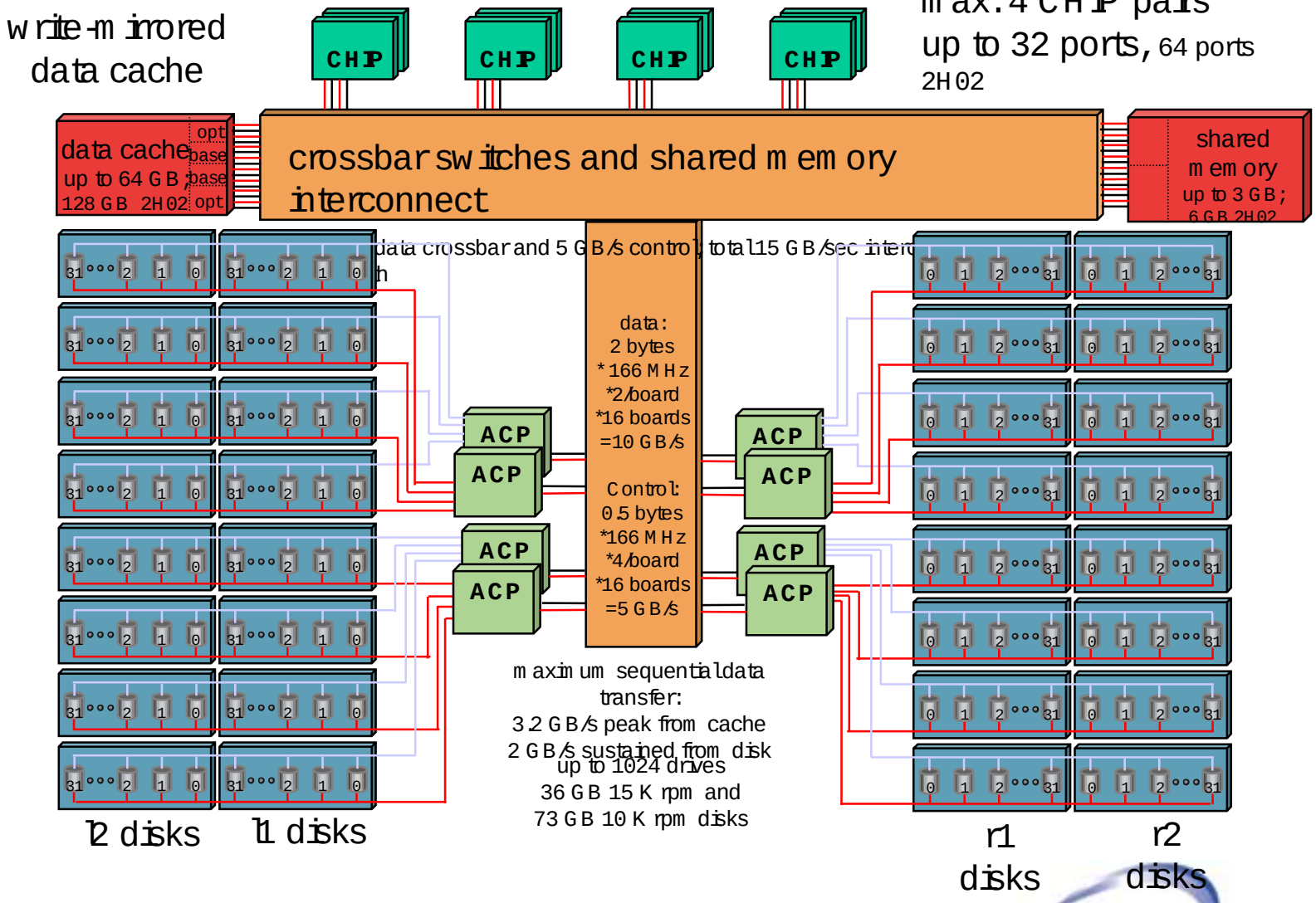
# Performance Improvement through Effective Disk Allocation and Logical Volume Striping - 258


| Danny B. Gross | Chris D. Roberson |
| --- | --- |
| Member of Technical Staff | Storage Solutions Architect Lead |
| Motorola, Inc | Hewlett-Packard |
| 6501 William Cannon West | 12401 Research Blvd, Suite 200 |
| Austin, TX 78735 | Austin, TX 78759 |
| (512) 895-4825 | (512) 257-5721 |
| danny.gross@motorola.com | chris.roberson@hp.com |

# Agenda

- Review XP Disk Array Architect
- Define Successful Storage Management Practices
- Perform Monitor of Storage Management Environment
- Summary

# xp1024 architecture



# Successful Storage Management and Layout

- Identify all storage requirements
- Create a data layout definition
- Map a logical disk layout
- Implement data layout

# Storage Requirements Map

- Identify all applications supported by storage environment
- Classify service level requirements
  - Availability
  - Performance
  - Capacity

| Equiva Storage Requirements | | | | | |
|---|---|---|---|---|---|
| **Application** | **Hosts** | **Availability** | **Performance** | **Capacity** | **Notes** |
| SAP | | | | | |
| | mapdb00p | Clustered | Random | 4100 | |
| | mapdb02p | Clustered | Random | 0 | |
| | mapin00p | | Random | 400 | |
| | ndceqbk1 | | Sequential | 4100 | |
| Data Warehouse | | | | | |
| | maw db01p | | Sequential | 1800 | |
| | lumdb05p | | Sequential | 500 | |
| | tstrept | | Sequential | 500 | |
| | ndceqbk2 | | Sequential | 1800 | |
| | | | | | |

# Data Definition Layout

The definition provides the implementation plan for all hosts attached to a storage network or network-attached storage

- mountpoints
- volume management information
- disk device requirements
- capacity used and free
- ldevs needed for volume groups

# Data Definition Layout “Greenfield” Example

| Force Production Existing Layout | | | | Force Production Migrating Layout | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| VG | LV | Mount Point | Requested LVM Total (Mbytes) | VG | LV | Mount Point | Open-E LDEVs | VG Total (Mbytes) | LVM Total (Mbytes) |
| vg03 | lvol1 | /u003 | 5000 | vg001 | lvol1 | /u003 | 4 | 28112 | 7028 |
| vg09 | lvol1 | /u009 | 2000 | | lvol2 | /u009 | | | 7028 |
| vg04 | lvol1 | /u004 | 5000 | | lvol3 | /u004 | | | 7028 |
| vg02 | lvol1 | /u002 | 7000 | vg002 | lvol1 | /u002 | 4 | 28112 | 7028 |
| vg10 | lvol1 | /u010 | 2000 | | lvol2 | /u010 | | | 7028 |
| vg12 | lvol1 | /u012 | 5000 | | lvol3 | /u012 | | | 7028 |
| vg01 | lvol1 | /u001 | 18000 | vg003 | lvol1 | /u001 | 12 | 84336 | 28112 |
| vg05 | lvol1 | /u005 | 9000 | | lvol2 | /u005 | | | 28112 |
| vg08 | lvol1 | /u008 | 8000 | | lvol3 | /u008 | | | 28112 |
| vg07 | lvol1 | /u007 | 6000 | vg004 | lvol1 | /u007 | 12 | 84336 | 28112 |
| vg06 | lvol1 | /u006 | 7000 | | lvol2 | /u006 | | | 28112 |
| vg11 | lvol1 | /u011 | 12000 | | lvol3 | /u011 | | | 28112 |
| | | | | | | | | | |
| Total: | | | 86000 | | | | | | 210840 |

| Force Non-Production Existing Layout | | | | Force Non-Production Migrating Layout | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| VG | LV | Mount Point | Requested LVM Total (Mbytes) | VG | LV | Mount Point | Open-E LDEVs | VG Total (Mbytes) | LVM Total (Mbytes) |
| vg09 | lvol1 | /u109 | 6000 | vg101 | | /u109 | 4 | 28112 | 7028 |
| vg03 | lvol1 | /u103 | 15000 | | | /u103 | | | 14056 |
| vg04 | lvol1 | /u104 | 15000 | vg102 | | /u104 | 4 | 28112 | 14056 |
| vg12 | lvol1 | /u112 | 15000 | | | /u112 | | | 14056 |
| vg02 | lvol1 | /u102 | 21000 | vg103 | | /u102 | 4 | 28112 | 21084 |
| vg10 | lvol1 | /u110 | 6000 | | | /u110 | | | 7028 |
| vg05 | lvol1 | /u105 | 27000 | vg104 | | /u105 | 12 | 84336 | 42168 |
| vg08 | lvol1 | /u108 | 24000 | | | /u108 | | | 42168 |
| vg07 | lvol1 | /u107 | 18000 | vg105 | | /u107 | 12 | 84336 | 42168 |
| vg06 | lvol1 | /u106 | 21000 | | | /u106 | | | 42168 |
| vg11 | lvol1 | /u111 | 36000 | vg106 | | /u111 | 12 | 84336 | 28112 |
| vg01 | lvol1 | /u101 | 54000 | | | /u101 | | | 56224 |
| | | | | | | | | | |
| Total: | | | 258000 | | | | | | 330316 |

| Yavin Production Existing Layout | | | | Yavin Production Migrating Layout | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| VG | LV | Mount Point | Requested LVM Total (Mbytes) | VG | LV | Mount Point | Open-E LDEVs | VG Total (Mbytes) | LVM Total (Mbytes) |
| vg01 | lvol1 | /u201 | 13 | vg200 | lvol1 | /u201 | 3 | 21084 | 21084 |
| vg02 | lvol1 | /u201 | 13 | vg201 | lvol1 | /u202 | 6 | 42168 | 42168 |
| | | | | | | | | | |
| Total: | | | 26 | | | | | | 63252 |

| Endor Production Existing Layout | | | | Endor Production Migrating Layout | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| VG | LV | Mount Point | Requested LVM Total (Mbytes) | VG | LV | Mount Point | Open-E LDEVs | VG Total (Mbytes) | LVM Total (Mbytes) |
| vg01 | lvol1 | /u201 | 13 | vg300 | lvol1 | /u301 | 3 | 21084 | 21084 |
| vg02 | lvol1 | /u201 | 13 | vg301 | lvol1 | /u302 | 6 | 42168 | 42168 |
| | | | | | | | | | |
| Total: | | | 26 | | | | | | 63252 |

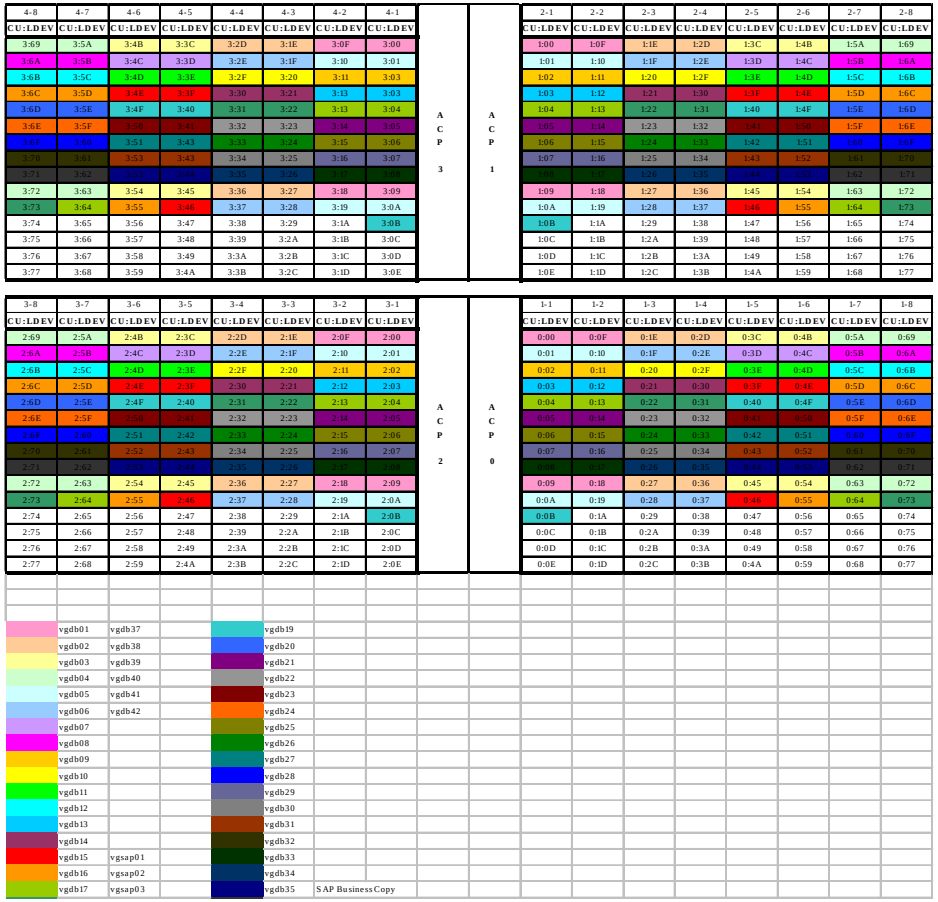
- Application team identified:
  - Database
  - Forms
  - Applications
- Infrastructure team identified:
  - Availability
  - Capacity
  - Backup Strategy

# Data Definition Layout “Retrofitted Environment

| Existing Layout | | | | | | | Migrating Layout | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VG | LV | Mount Point | BDF Total (Mbytes) | BDF Used (Mbytes) | BDF Free (Mbytes) | LVM Total (Mbytes) | VG | LV | Mount Point | Open-E LDEVs | VG Total (Mbytes) | LVM Total (Mbytes) | Datafile Migration (Mbytes) |
| vg_ixos | itadm | /itadm | 4096 | 3590 | 474 | 4000 | vg_ixos | itadm | /itadm | 4 | 55488 | 5000 | |
| vg_ixos | itparch | /archive/ITP | 5128 | 102 | 4712 | 5008 | | itparch | /archive/ITP | | | 5000 | |
| vg_ixos | itporadata1 | /oradata1/ITP | 5128 | 3202 | 1806 | 5008 | | itporadata1 | /oradata1/ITP | | | 5000 | |
| vg_ixos | itporadata2 | /oradata2/ITP | 5128 | 2587 | 2382 | 5008 | | itporadata2 | /oradata2/ITP | | | 5000 | |
| vg_ixos | itporadata3 | /oradata3/ITP | 5128 | 2613 | 2358 | 5008 | | itporadata3 | /oradata3/ITP | | | 5000 | |
| vg_ixos | itporadata4 | /oradata4/ITP | 3080 | 2649 | 404 | 3008 | | itporadata4 | /oradata4/ITP | | | 5000 | |
| vg_ixos | taxfirm | /export/prddata/multi/taxfirm | 20480 | 4 | 20157 | 20000 | | taxfirm | /export/prddata/multi/taxfirm | | | 20000 | |
| vg_nfs_infr | lv_controlm | /opt/controlm | 2572 | 2115 | 429 | 2512 | vg_nfs_infr | lv_controlm | /opt/controlm | 4 | 55488 | 2500 | |
| vg_nfs_infr | lv_ctmagent | /opt/ctmagent | 262 | 79 | 174 | 256 | | lv_ctmagent | /opt/ctmagent | | | 2500 | |
| vg_nfs_infr | lv_ecs | /opt/ecs | 2048 | 1070 | 917 | 2000 | | lv_ecs | /opt/ecs | | | 2500 | |
| vg_nfs_infr | lv_perfdata | /export/perfdata | 3080 | 2 | 2886 | 3008 | | lv_perfdata | /export/perfdata | | | 5000 | |
| vg_nfs_infr | lv_prddata | /export/prddata | 22561 | 19606 | 2909 | 22032 | | lv_prddata | /export/prddata | | | 25000 | |
| vg_nfs_infr | lv_prddata_data | /export/prddata/multi/datawh/arch | 10240 | 6854 | 3280 | 10000 | | lv_prddata_datawh | /export/prddata/multi/datawh/arch | | | 10000 | |
| vg_nfs_infr | lv_prod_doc | /export/jobsched | 49 | 8 | 39 | 48 | | lv_prod_doc | /export/jobsched | | | 2500 | |
| vg_nfs_infr | lv_scripts | /export/scripts | 164 | 73 | 85 | 160 | | lv_scripts | /export/scripts | | | 2500 | |
| vg_oracle_infr | lv_apps | /export/mapin01p | 9011 | 7890 | 1088 | 8800 | vg_oracle_infr | lv_apps | /export/mapin01p | 4 | 55488 | 10000 | |
| vg_oracle_infr | lv_arch | /export/mapin01p1/arch | 1032 | 1 | 966 | 1008 | | lv_arch | /export/mapin01p1/arch | | | 2500 | |
| vg_oracle_infr | lv_controlm_bk | /controlm_backup | 4096 | 561 | 3314 | 4000 | | lv_controlm_bk | /controlm_backup | | | 5000 | |
| vg_oracle_infr | lv_oracle | /oracleINFR | 410 | 1 | 383 | 400 | | lv_oracle | /oracleINFR | | | 2500 | |
| vg_oracle_infr | lv_oracle_8 | /oracle | 1540 | 6 | 1438 | 1504 | | lv_oracle_8 | /oracle | | | 2500 | |
| vg_oracle_infr | lv_oracle1 | /export/mapin01p1 | 1573 | 533 | 975 | 1536 | | lv_oracle1 | /export/mapin01p1 | | | 2500 | |
| vg_oracle_infr | lv_oracle2 | /export/mapin01p2 | 1573 | 416 | 1085 | 1536 | | lv_oracle2 | /export/mapin01p2 | | | 2500 | |
| vg_oracle_infr | lv_oracle3 | /export/mapin01p3 | 1573 | 1235 | 317 | 1536 | | lv_oracle3 | /export/mapin01p3 | | | 2500 | |
| vg_oracle_infr | lv_oracle4 | /export/mapin01p4 | 1540 | 678 | 808 | 1504 | | lv_oracle4 | /export/mapin01p4 | | | 2500 | |
| | | | | | | | | | | | | | |
| Total: | | | | | | 108880 | | | | | | 135000 | |

- Inventory existing environment
  - All hosts
  - All storage
  - Backup requirements
- Morph existing layout into redeployed layout

# Logical Disk Map




- Determine the quantity of physical volumes
- Color code unallocated CU:LDEV in 1st quadrant
- Extend color code to all CU:LDEVs in remaining quadrant
- Ensure all CU:LDEVs are contains within identical PVG

# Implementing the disk infrastructure

- “Greenfield”
  - Create fully new disk infrastructure
  - Copy or restore data to the replacement disks
- Relocate logical volume on a disk-for-disk basis
  - “pvmove” to relocate the extents of a logical volume from one disk to another
  - recover the obsolete disks by reducing from the volume group

# Implementing the disk infrastructure (Cont.)

- Re-stripe using mirrors in the target location
  - Organize the target disks into “physical volume groups
  - Insure that the logical volume is set to distributed and striped.
  - Create a striped mirror into the target physical volume group
  - Reduce the mirror from the obsolete location
  - Remove the obsolete disks from the volume group

# Measuring Performance of the Data Definition Layout

- UNIX I/O statistics Reporter “iostat”
  - Provides real-time disk latencies
  - Calculates real-time disk throughput
- UNIX Systems Activity Reporter “sar”
  - Provides more granular data than “iostat”
  - Allows generalized trending over short duration
- HP Perfview
  - Provides system analysis over long duration
  - Data collection performed over 5 minute interval

# “iostat”

| device | bps | sps | msps | | |
|---|---|---|---|---|---|
| c1t6d0 | 127 | 28.5 | 1.0 | vg00 | |
| c2t6d0 | 118 | 24.5 | 1.0 | vg00 | |
| c8t9d5 | 1025 | 44.4 | 1.0 | | /dev/vg_infmx/lvol |
| c6t9d5 | 125 | 4.8 | 1.0 | | /dev/vg_infmx/lvol |
| c5t9d5 | 125 | 4.8 | 1.0 | | /dev/vg_infmx/lvol |
| c8t9d6 | 136 | 5.2 | 1.0 | | /dev/vg_infmx/lvol |
| c6t9d6 | 2928 | 210.4 | 1.0 | | /dev/vg_infmx/lvol |
| c5t9d6 | 362 | 31.9 | 1.0 | | /dev/vg_infmx/lvol |
| c8t10d4 | 2792 | 304.6 | 1.0 | | /dev/vg_infmx/lvol |
| c6t10d4 | 1103 | 146.0 | 1.0 | | /dev/vg_infmx/lvol |
| c5t10d4 | 135 | 6.0 | 1.0 | | /dev/vg_infmx/lvol |
| c33t1d6 | 948 | 8.0 | 1.0 | | /dev/vg_infmx/lvol |
| c35t4d1 | 948 | 8.0 | 1.0 | | |
| c29t6d0 | 948 | 8.0 | 1.0 | | |
| c31t10d2 | 948 | 8.0 | 1.0 | | |
| c26t1d6 | 948 | 8.0 | 1.0 | | |
| c33t6d0 | 948 | 8.0 | 1.0 | | |
| c35t8d3 | 948 | 8.0 | 1.0 | | |
| c29t10d2 | 948 | 8.0 | 1.0 | | |
| c31t1d6 | 948 | 8.0 | 1.0 | | |
| c26t6d0 | 948 | 8.0 | 1.0 | | |
| c33t10d2 | 948 | 8.0 | 1.0 | | |
| c35t12d5 | 1205 | 9.0 | 1.0 | | |

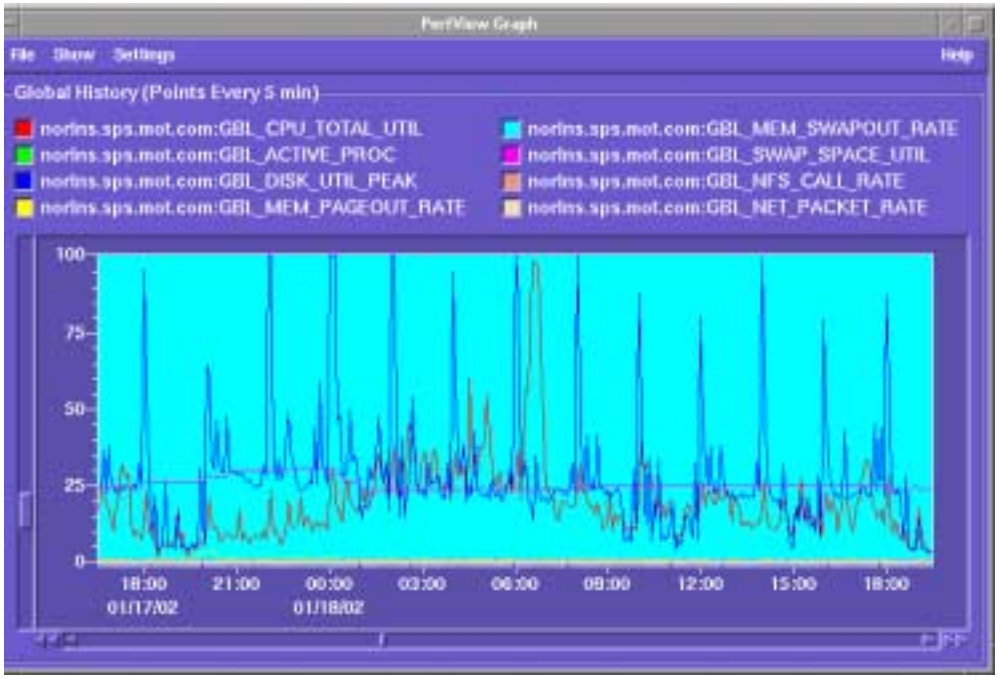
- /dev/vg_infmx/lvol1 is a distributed but unstriped volume
- /dev/vg_infmx/lvol2 is a similar volume, built on four disks
- High amount of seeks on the busier disks
- I/O evenly distributed across all twelve of the mirror disks

# “sar”

| HP-UX | system1 | B.11.00 | A | 9000/800 | 2/4/2002 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | ms | ms | ms | | |
| 9:39:26 | device | %busy | avque | r+w/s | blks/s | await | avserv | Totwait | AveTotal | Aver+w/s | Aveblks/s | AveBusy |
| | | | | | | | | | wait | | | |
| Average | c1t6d0 | 5.64 | 0.56 | 9 | 71 | 4.95 | 9.62 | 14.57 | | | | |
| Average | c2t6d0 | 4.23 | 0.56 | 8 | 65 | 4.95 | 7.7 | 12.65 | | | | |
| Average | c17t10d7 | 35.67 | 0.59 | 223 | 2301 | 5.23 | 2.12 | 7.35 | | | | |
| Average | c17t2d7 | 27.59 | 1.9 | 70 | 3164 | 7.49 | 13 | 20.49 | | | | |
| Average | c18t2d7 | 23.88 | 1.85 | 56 | 3066 | 7.53 | 14.2 | 21.73 | | | | |
| Average | c10t2d7 | 22.43 | 1.75 | 51 | 3046 | 7.31 | 13.8 | 21.11 | | | | |
| Average | c10t10d7 | 6.47 | 1.24 | 35 | 1804 | 6.77 | 5.15 | 11.92 | | | | |
| Average | c18t10d7 | 6.26 | 0.94 | 32 | 1420 | 5.95 | 4.01 | 9.96 | | | | |
| Average | c10t5d4 | 0.23 | 0.5 | 1 | 10 | 3.8 | 4.32 | 8.12 | 14.38 | 66.86 | 2115.86 | 17.50 |
| Average | c31t1d1 | 22.45 | 0.52 | 211 | 1680 | 5.08 | 1.27 | 6.35 | | | | |
| Average | c33t1d1 | 15.41 | 0.97 | 53 | 2226 | 5.95 | 8.17 | 14.12 | | | | |
| Average | c37t12d0 | 14.42 | 1.93 | 35 | 2443 | 7.44 | 12.89 | 20.33 | | | | |
| Average | c29t5d3 | 13.2 | 1.38 | 31 | 2236 | 6.53 | 13.08 | 19.61 | | | | |
| Average | c35t9d5 | 12.88 | 1.41 | 32 | 2213 | 6.77 | 12.3 | 19.07 | | | | |
| Average | c35t11d4 | 4.05 | 1.33 | 18 | 1317 | 8.25 | 7.71 | 15.96 | | | | |
| Average | c29t7d2 | 3.66 | 0.86 | 19 | 1541 | 5.73 | 4.98 | 10.71 | | | | |
| Average | c33t3d0 | 3.05 | 0.83 | 16 | 855 | 5.92 | 4.62 | 10.54 | | | | |
| Average | c33t4d7 | 0.05 | 0.5 | 0 | 5 | 3.34 | 3.9 | 7.24 | 13.77 | 46.11 | 1612.89 | 9.91 |
| | | | | | | | | | | | | |
| | | | | | | | Summary: | | 4.26% | 31.03% | 23.77% | 43.40% |

- busy is reduced by 43.4% on average
- r+w/s is reduced by 31.0% on average
- Total Wait time is reduced by 4.3% on average
- Average blks/s is reduced by 23.8% on average

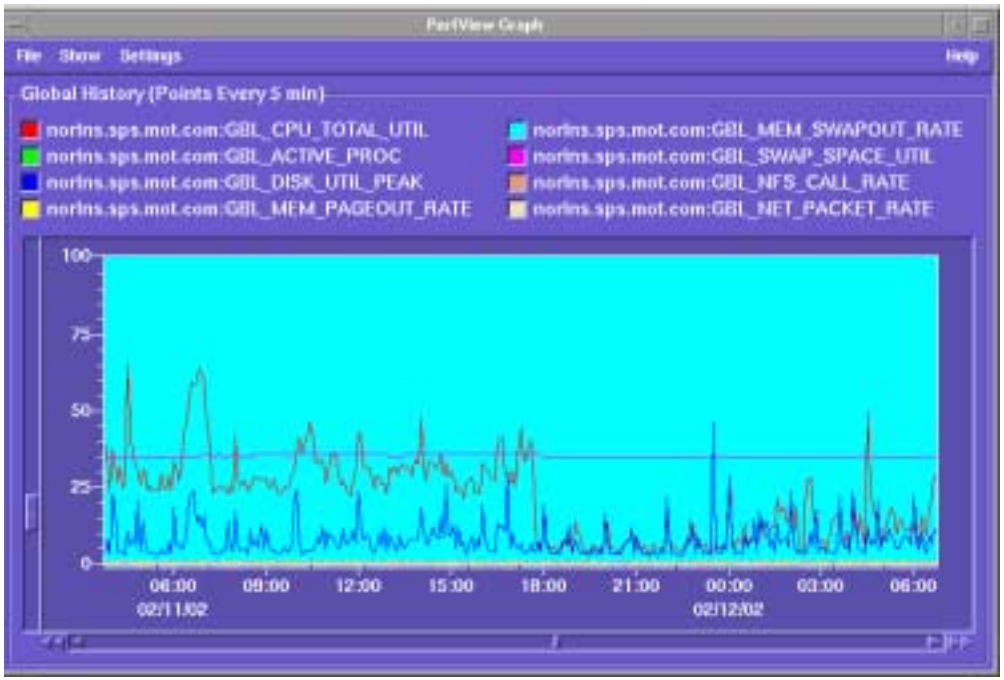
# PerfView Pre-migration




- CPU load approximately 15 – 25 %
- Disk utilization at 25%, peaking frequently at 100% .

This data was fairly representative of day-by-day operations of the system over several months.

# Perfview Post-migration




- Disk Utilization averages 5%, peaking at 45%.
- CPU Utilization averages 15%

# Summary

- Increase storage utilization by deploying sound management practices
- Design phase factors out implementation risk
- Several techniques exist for implementation and deployment of this strategy
- Performance tools assist with evaluating the environment change