

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

Danny B. Gross  
Member of Technical Staff  
Motorola, Inc  
6501 William Cannon West  
Austin, TX 78735  
(512) 895-4825  
danny.gross@motorola.com

Chris D. Roberson  
Storage Solutions Architect Lead  
Hewlett-Packard  
12401 Research Blvd, Suite 200  
Austin, TX 78759  
(512) 257-5721  
chris.roberson@hp.com

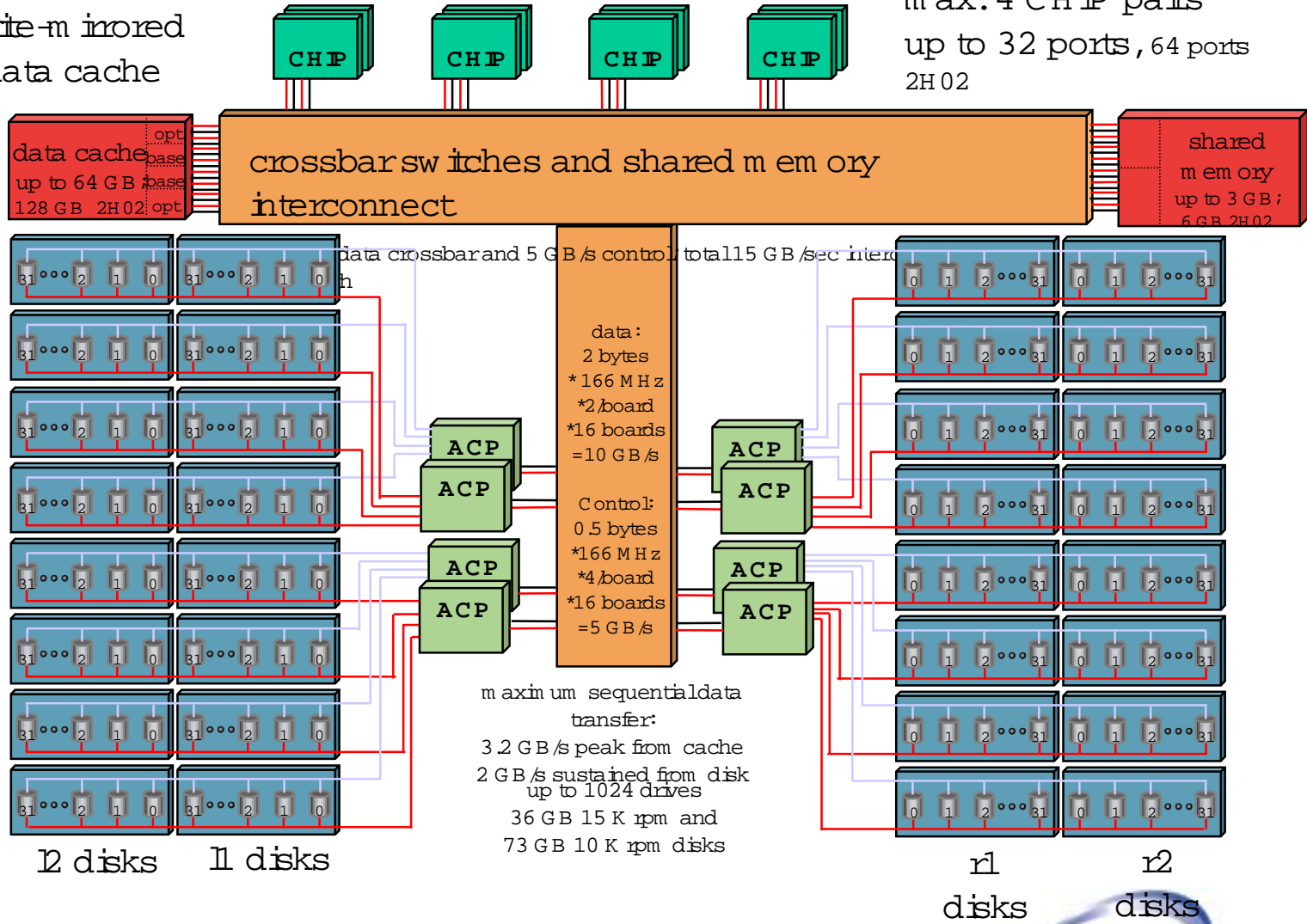
# Agenda

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

# xp1024 architecture

write mirrored  
data cache

max. 4 CHIP pairs  
up to 32 ports, 64 ports  
2H02



# 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	lv01	/u003	5000	vg001	lv01	/u003	4	28112	7028
vg09	lv01	/u009	2000		lv02	/u009			7028
vg04	lv01	/u004	5000		lv03	/u004			7028
vg02	lv01	/u002	7000	vg002	lv01	/u002	4	28112	7028
vg10	lv01	/u010	2000		lv02	/u010			7028
vg12	lv01	/u012	5000		lv03	/u012			7028
vg01	lv01	/u001	8000	vg003	lv01	/u001	12	84336	28112
vg05	lv01	/u005	9000		lv02	/u005			28112
vg08	lv01	/u008	8000		lv03	/u008			28112
vg07	lv01	/u007	6000	vg004	lv01	/u007	12	84336	28112
vg06	lv01	/u006	7000		lv02	/u006			28112
vg11	lv01	/u011	12000		lv03	/u011			28112
<b>Total:</b>			<b>86000</b>						<b>210840</b>

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	lv01	/u109	6000	vg101	lv01	/u109	4	28112	7028
vg03	lv01	/u103	15000		lv03				14056
vg04	lv01	/u104	15000	vg102	lv02	/u104	4	28112	14056
vg12	lv01	/u112	15000		lv112				14056
vg02	lv01	/u102	23000	vg103	lv03	/u102	4	28112	21084
vg10	lv01	/u110	6000		lv110				7028
vg05	lv01	/u105	27000	vg104	lv05	/u105	12	84336	42168
vg08	lv01	/u108	24000		lv108				42168
vg07	lv01	/u107	8000	vg105	lv05	/u107	12	84336	42168
vg06	lv01	/u106	23000		lv106				42168
vg11	lv01	/u111	36000	vg106	lv06	/u111	12	84336	28112
vg01	lv01	/u101	34000		lv101				56224
<b>Total:</b>			<b>258000</b>						<b>330316</b>

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	lv01	/u201	13	vg200	lv01	/u201	3	21084	21084
vg02	lv01	/u201	13	vg201	lv01	/u202	6	42168	42168
<b>Total:</b>			<b>26</b>						<b>63252</b>

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	lv01	/u201	13	vg300	lv01	/u301	3	21084	21084
vg02	lv01	/u201	13	vg301	lv01	/u302	6	42168	42168
<b>Total:</b>			<b>26</b>						<b>63252</b>

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

# Data Definition Layout “Retrofitted Environment

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

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_100s	indm	/indm	4096	3990	474	4000	vg_100s	indm	/indm	4	55488	5000	
vg_100s	nparch	/archive/1TP	5128	802	4712	5008	vg_100s	nparch	/archive/1TP			5000	
vg_100s	nporadata1	/oradata1/1TP	5128	3202	1806	5008	vg_100s	nporadata1	/oradata1/1TP			5000	
vg_100s	nporadata2	/oradata2/1TP	5128	2587	2382	5008	vg_100s	nporadata2	/oradata2/1TP			5000	
vg_100s	nporadata3	/oradata3/1TP	5128	2613	2358	5008	vg_100s	nporadata3	/oradata3/1TP			5000	
vg_100s	nporadata4	/oradata4/1TP	3080	2649	404	3008	vg_100s	nporadata4	/oradata4/1TP			5000	
vg_100s	taxfirm	/export/prddata/multi/taxfirm	20480	4	20157	20000	vg_100s	taxfirm	/export/prddata/multi/taxfirm			20000	
vg_nfs_infr	lv_controlm	/opt/controlm	2572	218	429	2512	vg_nfs_infr	lv_controlm	/opt/controlm	4	55488	2500	
vg_nfs_infr	lv_etmagent	/opt/etmagent	262	79	174	256	vg_nfs_infr	lv_etmagent	/opt/etmagent			2500	
vg_nfs_infr	lv_ees	/opt/ees	2048	1070	977	2000	vg_nfs_infr	lv_ees	/opt/ees			2500	
vg_nfs_infr	lv_perfddata	/export/perfddata	3080	2	2886	3008	vg_nfs_infr	lv_perfddata	/export/perfddata			2500	
vg_nfs_infr	lv_prddata	/export/prddata	22561	19606	2909	22022	vg_nfs_infr	lv_prddata	/export/prddata			25000	
vg_nfs_infr	lv_prddata_data	/export/prddata/multi/datawh/arch	10240	6854	3280	10000	vg_nfs_infr	lv_prddata_datawh	/export/prddata/multi/datawh/arch			10000	
vg_nfs_infr	lv_prod_doc	/export/jobshed	49	8	39	48	vg_nfs_infr	lv_prod_doc	/export/jobshed			2500	
vg_nfs_infr	lv_scripts	/export/scripts	164	73	85	160	vg_nfs_infr	lv_scripts	/export/scripts			2500	
vg_oracle_infr	lv_apps	/export/mapin0 lp	9011	7890	1088	8800	vg_oracle_infr	lv_apps	/export/mapin0 lp	4	55488	10000	
vg_oracle_infr	lv_arch	/export/mapin0 lp/arch	1032	1	966	1008	vg_oracle_infr	lv_arch	/export/mapin0 lp/arch			2500	
vg_oracle_infr	lv_controlm_bk	/controlm_backup	4096	561	3334	4000	vg_oracle_infr	lv_controlm_bk	/controlm_backup			5000	
vg_oracle_infr	lv_oracle	/oracleBFR	410	1	383	400	vg_oracle_infr	lv_oracle	/oracleBFR			2500	
vg_oracle_infr	lv_oracle_8	/oracle	1540	6	1438	1504	vg_oracle_infr	lv_oracle_8	/oracle			2500	
vg_oracle_infr	lv_oracle1	/export/mapin0 lp1	1573	533	975	1536	vg_oracle_infr	lv_oracle1	/export/mapin0 lp1			2500	
vg_oracle_infr	lv_oracle2	/export/mapin0 lp2	1573	416	1085	1536	vg_oracle_infr	lv_oracle2	/export/mapin0 lp2			2500	
vg_oracle_infr	lv_oracle3	/export/mapin0 lp3	1573	1235	317	1536	vg_oracle_infr	lv_oracle3	/export/mapin0 lp3			2500	
vg_oracle_infr	lv_oracle4	/export/mapin0 lp4	1540	678	808	1504	vg_oracle_infr	lv_oracle4	/export/mapin0 lp4			2500	
<b>Total:</b>						<b>108880</b>						<b>135000</b>	



# Logical Disk Map

4:8	4:7	4:6	4:5	4:4	4:3	4:2	4:1			2:1	2:2	2:3	2:4	2:5	2:6	2:7	2:8		
CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	A C P 3	A C P 1	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV		
3:69	3:5A	3:4B	3:3C	3:2D	3:1E	3:0F	3:00			1:00	1:0F	1:1E	1:2D	1:3C	1:4B	1:5A	1:69		
3:6A	3:5B	3:4C	3:3D	3:2E	3:1F	3:10	3:01			1:01	1:10	1:1F	1:2E	1:3D	1:4C	1:5B	1:6A		
3:6B	3:5C	3:4D	3:3E	3:2F	3:20	3:11	3:03			1:02	1:11	1:20	1:2F	1:3E	1:4D	1:5C	1:6B		
3:6C	3:5D	3:4E	3:3F	3:30	3:21	3:13	3:05			1:03	1:12	1:21	1:30	1:3F	1:4E	1:5D	1:6C		
3:6D	3:5E	3:4F	3:40	3:31	3:22	3:15	3:04			1:04	1:13	1:22	1:31	1:40	1:4F	1:5E	1:6D		
3:6E	3:5F	3:40	3:41	3:32	3:23	3:14	3:05			1:05	1:14	1:23	1:32	1:41	1:50	1:5F	1:6E		
3:6F	3:50	3:51	3:43	3:34	3:25	3:16	3:06			1:06	1:15	1:24	1:33	1:42	1:51	1:60	3:07		
3:70	3:42	3:43	3:44	3:35	3:26	3:17	3:07			1:07	1:16	1:25	1:34	1:43	1:52	1:61	3:08		
3:72	3:43	3:44	3:45	3:36	3:27	3:18	3:09			1:09	1:18	1:27	1:36	1:45	1:54	1:63	3:09		
3:73	3:44	3:45	3:46	3:37	3:28	3:19	3:0A			1:0A	1:19	1:28	1:37	1:46	1:55	1:64	3:0A		
3:74	3:45	3:46	3:47	3:38	3:29	3:1A	3:0B			1:0B	1:1A	1:29	1:38	1:47	1:56	1:65	3:0B		
3:75	3:46	3:47	3:48	3:39	3:2A	3:1B	3:0C			1:0C	1:1B	1:2A	1:39	1:48	1:57	1:66	3:0C		
3:76	3:47	3:48	3:49	3:3A	3:2B	3:1C	3:0D			1:0D	1:1C	1:2B	1:3A	1:49	1:58	1:67	3:0D		
3:77	3:48	3:49	3:4A	3:3B	3:2C	3:1D	3:0E			1:0E	1:1D	1:2C	1:3B	1:4A	1:59	1:68	3:0E		
3:8	3:7	3:6	3:5	3:4	3:3	3:2	3:1					1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8
CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV			A C P 2	A C P 0	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV	CU:LDEV
2:69	2:5A	2:4B	2:3C	2:2D	2:1E	2:0F	2:00	0:00	0:0F			0:1E	0:2D	0:3C	0:4B	0:5A	0:69		
2:6A	2:5B	2:4C	2:3D	2:2E	2:1F	2:10	2:01	0:01	0:10			0:1F	0:2E	0:3D	0:4C	0:5B	0:6A		
2:6B	2:5C	2:4D	2:3E	2:2F	2:20	2:11	2:03	0:02	0:11			0:20	0:2F	0:3E	0:4D	0:5C	0:6B		
2:6C	2:5D	2:4E	2:3F	2:30	2:21	2:13	2:05	0:03	0:12			0:21	0:30	0:3F	0:4E	0:5D	0:6C		
2:6D	2:5E	2:4F	2:40	2:31	2:22	2:15	2:04	0:04	0:13			0:22	0:31	0:40	0:4F	0:5E	0:6D		
2:6E	2:5F	2:40	2:41	2:32	2:23	2:14	2:05	0:05	0:14			0:23	0:32	0:41	0:50	0:5F	0:6E		
2:6F	2:50	2:51	2:43	2:34	2:25	2:16	2:06	0:06	0:15			0:24	0:33	0:42	0:51	0:60	2:07		
2:70	2:42	2:43	2:44	2:35	2:26	2:17	2:07	0:07	0:16			0:25	0:34	0:43	0:52	0:61	2:08		
2:72	2:43	2:44	2:45	2:36	2:27	2:18	2:09	0:09	0:18			0:27	0:36	0:45	0:54	0:63	2:09		
2:73	2:44	2:45	2:46	2:37	2:28	2:19	2:0A	0:0A	0:19			0:28	0:37	0:46	0:55	0:64	2:0A		
2:74	2:45	2:46	2:47	2:38	2:29	2:1A	2:0B	0:0B	0:1A			0:29	0:38	0:47	0:56	0:65	2:0B		
2:75	2:46	2:47	2:48	2:39	2:2A	2:1B	2:0C	0:0C	0:1B			0:2A	0:39	0:48	0:57	0:66	2:0C		
2:76	2:47	2:48	2:49	2:3A	2:2B	2:1C	2:0D	0:0D	0:1C			0:2B	0:3A	0:49	0:58	0:67	2:0D		
2:77	2:48	2:49	2:4A	2:3B	2:2C	2:1D	2:0E	0:0E	0:1D			0:2C	0:3B	0:4A	0:59	0:68	2:0E		
vgdb01	vgdb37						vgdb19												
vgdb02	vgdb38						vgdb20												
vgdb03	vgdb39						vgdb21												
vgdb04	vgdb40						vgdb22												
vgdb05	vgdb41						vgdb23												
vgdb06	vgdb42						vgdb24												
vgdb07							vgdb25												
vgdb08							vgdb26												
vgdb09							vgdb27												
vgdb10							vgdb28												
vgdb11							vgdb29												
vgdb12							vgdb30												
vgdb13							vgdb31												
vgdb14							vgdb32												
vgdb15	vgsap01						vgdb33												
vgdb16	vgsap02						vgdb34												
vgdb17	vgsap03						vgdb35												

- Determine the quantity of physical volumes
- Color code unallocated CU:LDEV in 1<sup>st</sup> 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	mmps	
c1t6d0	127	28.5	1.0	vg00
c2t6d0	118	24.5	1.0	vg00
c8t9d5	1025	44.4	1.0	/dev/vg_infmtx/lvol
c6t9d5	125	4.8	1.0	/dev/vg_infmtx/lvol
c5t9d5	125	4.8	1.0	/dev/vg_infmtx/lvol
c8t9d6	136	5.2	1.0	/dev/vg_infmtx/lvol
c6t9d6	2928	210.4	1.0	/dev/vg_infmtx/lvol
c5t9d6	362	31.9	1.0	/dev/vg_infmtx/lvol
c8t10d4	2792	304.6	1.0	/dev/vg_infmtx/lvol
c6t10d4	1103	146.0	1.0	/dev/vg_infmtx/lvol
c5t10d4	135	6.0	1.0	/dev/vg_infmtx/lvol
c33t1d6	948	8.0	1.0	/dev/vg_infmtx/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	

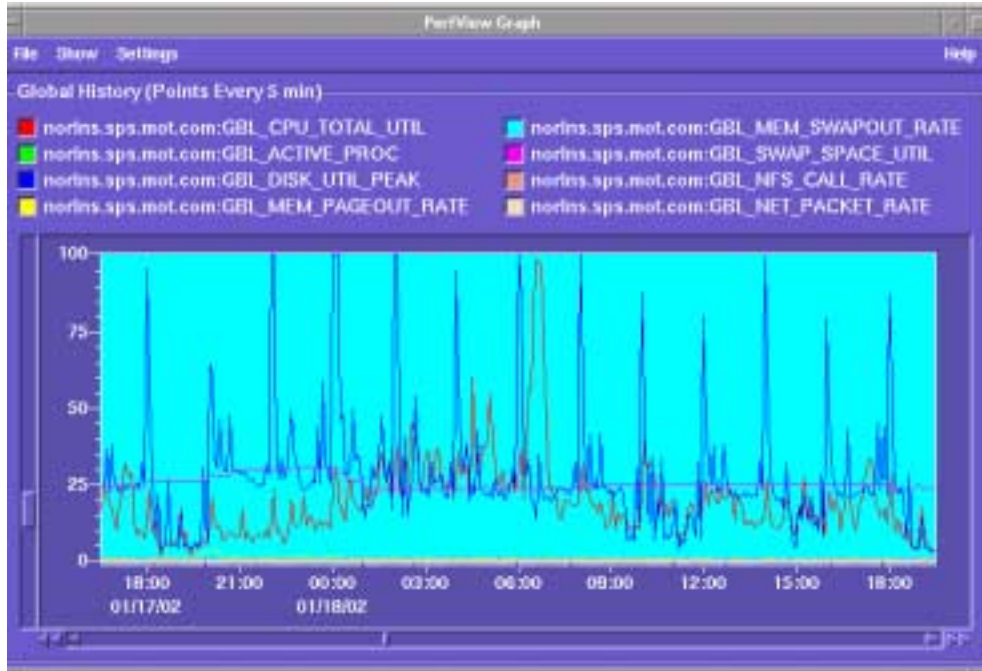
- /dev/vg\_infmtx/lvol1 is a distributed but unstriped volume
- /dev/vg\_infmtx/lvol2 is a similar volume, built on four disks
- High amount of seeks on the busier disks
- I/O evenly distributed across all twelve of the mirror disks

# “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

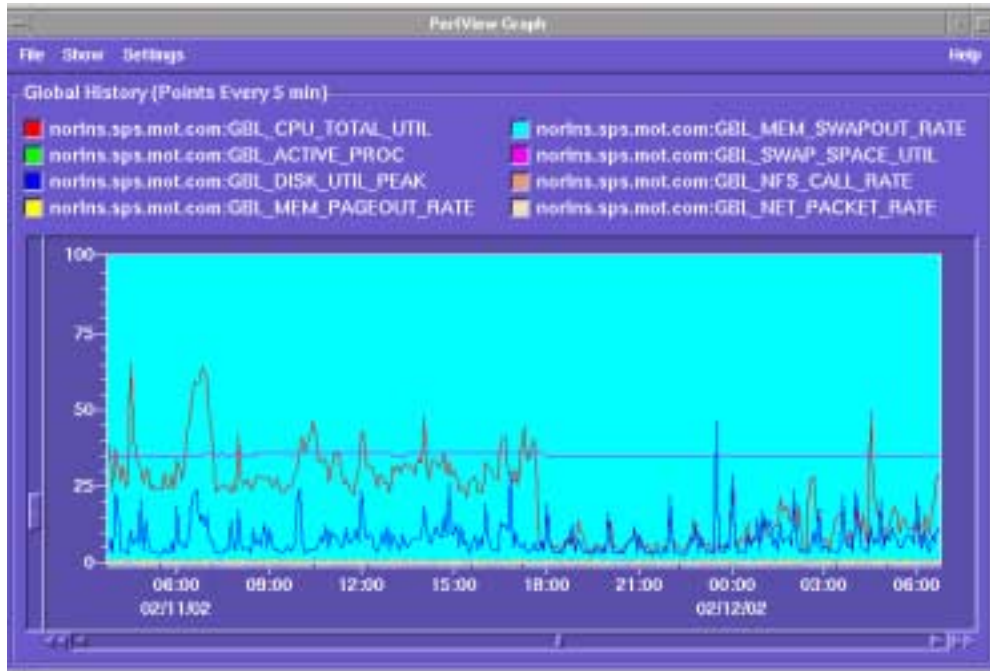
# PerfView Pre-migration



- 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