Best practices for Oracle 10g and EVA5000

Best practices for Oracle 10g and EVA5000

Introduction

The HP StorageWorks Enterprise Virtual Array 5000 is a fast, flexible and reliable storage array, supported by all major operating systems. It can be used efficiently for mission critical applications using Oracle databases.

This paper describes optimization techniques and best practices for configuring an Enterprise Virtual Array 5000 with Oracle 10g. Oracle Database Administrators, using an EVA5000 in an Oracle environment, can achieve higher performance and availability, along with simplified management capabilities they previously lacked. Other topics include the following:

- Optimal file layouts for an Oracle database
- Restriping and the associated effect on performance and availability
- Vsnaps and snapclones and how availability and throughput are affected
- A discussion of the Oracle Flexible Architechture to provide information on Oracle file distribution on available storage

For more information on EVA and general best practices for EVA see the HP StorageWorks Enterprise Virtual Array Configuration Best Practices white paper.

EVA5000

A storage system is composed of one HSV controller pair and either 2, 6, 8, or 12 disk-drive enclosures (Model 5214). Each Model 3220 Controller consists of two HSV110 controllers. The HP Enterprise Storage System only supports dual-controller, dual-loop operation. An Enterprise Storage System is monitored and managed centrally through the HSV Element Manager, a user-friendly graphical user interface (GUI) accessible through a standard internet browser. Here are some of the main features of the EVA

- **Hot-pluggable –** Including disk drives, the environmental monitoring unit (EMU), blowers, I/O modules, and controllers
- **OCP -** With LEDs, the LCD, and the pushbuttons are for initial setup and maintenance functions.
- **CRUs -** Customer replaceable units include the blowers, the batteries, the transceivers, and the cables.
- **Node WWN -** HSV Storage System operations require that each controller pair have a unique Node WWN (World Wide Name). This 16-character alphanumeric name identifies the system on the World Wide Web (WWW). HP assigns the WWN to each storage system prior to shipment.
- **Cache Battery -** When both battery assemblies are fully charged, they can provide power to the DIMMs for up to 96 hours.

The EVA5000 uses a virtualization technology in which physical disks are organized into disk groups. From a disk group, users can create virtual disks with a desired protection level and capacity. The virtual disks are distributed over the entire disk group, by spreading the data across the spindles, thereby avoiding hot spots on individual spindles. So effectively a RAID1 unit on an EVA is RAID1+0.

For example, if a user creates a disk group of 28 36G disks, the result is a total capacity of 1008G. From this raw pool, a user can create a RAID1 disk of 100G and 4 RAID5 disks of 160G each. These virtual disks will not only offer the RAID protection that was chosen, but IT will be spread across the 28 physical disks. For more information on the EVA, refer to the EVA Initial Setup User Guide.

Methodology

Testing

The testing was completed in 2 phases:
- Phase 1 was completed with the EVAs running VCS 3.0.1.0 with SSSU (See pg. 41 for more information on SSSU).
- Phase 2 was completed with EVAs running VCS 3.0.2.0 with Business Copy Software (BC). (See pg.41 for more information on BC.)

The Oracle workloads were run on the system and system performance graphs were captured to understand the resource utilization. Different EVA activities were then performed with the same Oracle workload running in order to study the effect of doing them together. System performance graphs were captured during each of these runs to understand the effect and also as proof that the system running the workloads was not impacted by the activities run on the EVA.

Phase 1: Using SSSU and VCS 3.0.1.0

Two Oracle 10g databases were created and laid out according to the Oracle recommendations on the EVA. The databases were run without archive logs data and data was loaded using a data generator and sql loader in a indirect mode (using redo logging) to simulate large number of inserts into the database. The test ran 16 streams in parallel to load 10,000,000 rows each (total of 160,000,000 rows) into each database.

After obtaining a baseline timing of the test, it was repeated with different EVA activities like creating snapclones, snapshots and restriping the disk group by adding disks. The EVA activities were performed using the SSSU utility and the HP SMA web interface.

Phase 2 : Using Business Copy and VCS 3.0.2.0

For the second part of the test, the workload was trebled by increasing the number of rows to 30,000,000 for each stream (for a total of 480,000,000) into each database. This was completed for the follow-on test, where the clones or snapshots were mounted on a different node and a backup was simulated by reading all the mounted devices and redirecting the output to a null device. The creation of the snapclones and snapshots were completed using the Business Copy software.

Configuration

Three Enterprise configurations were tested based on the points previously outlined, and results were provided for each.

Layout 1 : One large disk group (all logs and databases as well as snapclones in single disk group)

Layout 2 : Three pools of data, two small groups for logs (one for redo logs and one for archive logs) and one large group for database

Layout 3 : Two equal groups with logs and data split between both groups

These configurations were chosen based on input received from the HP sales field and broadly cover the typical configurations. The details of the virtual disks and their layout in the disk groups in each layout are described later in the document.

Exceptions

Some of the tests previously mentioned were not executed due to the following limitations:

- Snapshots were not created using SSSU on Layout 2 as the test required creating snapshots on a diskgroup other than the original disk and this is not allowed by the EVA.

- Snapshots or snapclones were not created using BC on Layout 2 for the same reasons.

- The Restriping operation was not repeated in phase II as adding disks to disk groups was not an SSSU / BC operation.

Hardware and Software Configuration

(1) HP AlphaServer, ES47 running Tru64 UNIX Version 5.1B with BL25
  − (4 1 GHz processors
  − (4) 2 Gb HBAs
  − Storage
    ▪ EVA 2C12D-C 60Hz Graphite (VCS 3.0.1.0 & 3.0.2.0)
  − SAN Components
    ▪ (4) Fibre Channel switch, 2/16
    ▪ SMA II (v3.3)
    ▪ SSSU build 7 for Tru64.
    ▪ Business Copy V 2.3

Oracle Release 10.1.0.2.0

Layout 1

For Layout 1, the EVA configuration consisted of one disk group with 34 Vdisks for Oracle files, redo logs and archive logs was created. The snapshots and snap clones were created in the same disk group.

Disk Group 1
DB1 control file, system datafile and data file 1
DB1 data file 2
DB1 data file 3
DB1 data file 4
DB1 data file 5
DB1 data file 6
DB1 data file 7
DB1 data file 8
DB1 data file 9
DB1 data file 10
DB1 redo log 1
DB1 redo log 2
DB1 redo log 3
DB1 redo log 4
DB1 archive log
DB1 temp file
DB1 undo file
DB2 control file, system datafile and data file 1
DB2 data file 2
DB2 data file 3
DB2 data file 4
DB2 data file 5
DB2 data file 6
DB2 data file 7
DB2 data file 8
DB2 data file 9
DB2 data file 10
DB2 redo log 1
DB2 redo log 2
DB2 redo log 3
DB2 redo log 4
DB2 archive log
DB2 temp file
DB2 undo file

Layout 2

For Layout 2, the disk groups and the virtual disks were created as shown in the following table. The intent was to separate the redo logs and archive logs from the data files. This kept the disks that get written into, in different groups. The snapshots and clones are created in a separate disk group for the same reasons.

Disk Group 1	Disk Group 2	Disk Group 3	Disk Group 4
DB1 control file, system datafile and data file 1	Snapshots & Clones	DB1 redo log 1	DB2 redo log 1
DB1 data file 2		DB1 redo log 2	DB2 redo log 2

DB1 data file 3		DB1 redo log 3	DB2 redo log 3
DB1 data file 4		DB1 redo log 4	DB2 redo log 4
DB1 data file 5		DB1 archive log	DB2 archive log
DB1 data file 6			
DB1 data file 7			
DB1 data file 8			
DB1 data file 9			
DB1 data file 10			
DB1 temp file			
DB1 undo file			
DB2 control file, system datafile and data file 1			
DB2 data file 2			
DB2 data file 3			
DB2 data file 4			
DB2 data file 5			
DB2 data file 6			
DB2 data file 7			
DB2 data file 8			
DB2 data file 9			
DB2 data file 10			
DB2 temp file			
DB2 undo file			

Layout 3

For Layout 3, the EVA configuration consisted of two disk groups (one for each database). The snap clones of the disks were created in the other disk group in phase 1 using SSSU. The snapshots were created in the same disk groups as the source disk. In phase 2, all the snap shots and clones were created in the same disk group as the source.

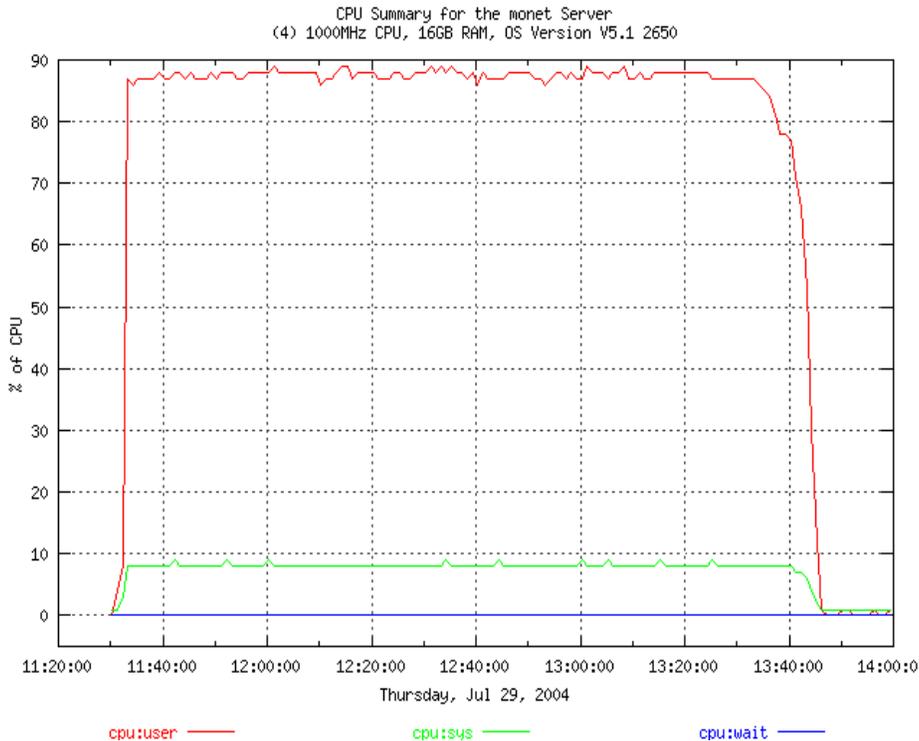| Disk Group 1	Disk Group 2
DB 1 control file, system data file, data file 1	DB2 control file, system data file and data file 1
DB1 data file 2	DB2 data file 2
DB1 data file 3	DB2 data file 3
DB1 data file 4	DB2 data file 4
DB1 data file 5	DB2 data file 5
DB1 data file 6	DB2 data file 6
DB1 data file 7	DB2 data file 7
DB1 data file 8	DB2 data file 8
DB1 data file 9	DB2 data file 9
DB1 data file 10	DB2 data file 10
DB1 temp file	DB2 temp file
DB1 undo file	DB2 undo file
DB1 redo log 1	DB2 redo log 1
DB1 redo log 2	DB2 redo log 2
DB1 redo log 3	DB2 redo log 3
DB1 redo log 4	DB2 redo log 4
DB1 archive log	DB2 archive log

Phase 1 Testing

For Phase 1 testing, the EVA was running VCS 3.0.1.0. The cloning and snapshotting was performed using SSSU.

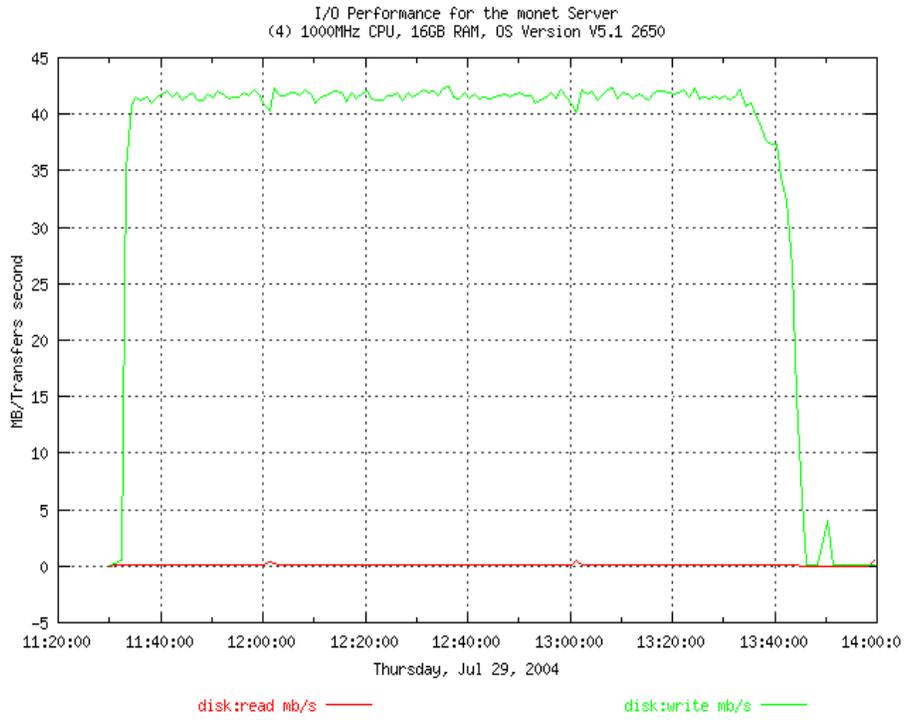## Baseline run without archive logs – Layout 1

In this test, the two databases were created in the first layout. Data was loaded into both databases with archiving turned off. The following tables list the runtime for these jobs.

	Start time	End time	Run time
Database A	11:32:14	13:43:52	02:11
Database B	11:32:14	13:44:55	02:12

The system CPU and I/O performance during this run was captured and is shown in the following graphs. The CPUs were completely utilized during the run and I/O (only writes) was steady over the duration of the test.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Best practices for Oracle 10g and EVA5000

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Thursday, Jul 29, 2004

disk:read mb/s ——   disk:write mb/s ——

Baseline run with archive logs – Layout 1

In this test, data was loaded into both databases with archiving turned on. This activity was run twice to verify that runtimes were the same. The following tables list the runtime for these jobs.
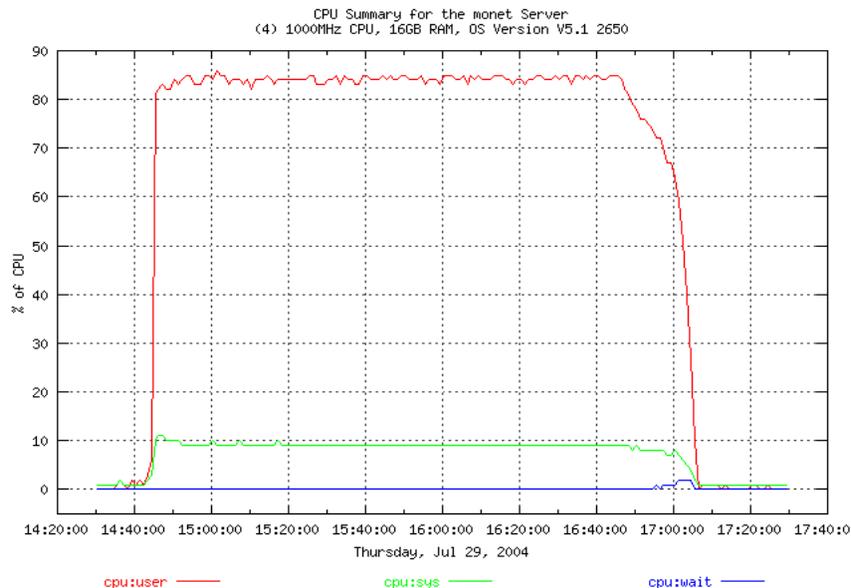
Run no. 1

	Start time	End time	Run time
Database A	14:44:22	17:05:17	02:21
Database B	14:44:23	17:04:01	02:20

Run no. 2

	Start time	End time	Run time
Database A	18:33:39	20:54:22	02:21
Database B	18:33:41	20:54:44	02:21

By ensuring that the archive logs are on separate disks and by distributing the redo logs, we ensure that no units are being read from or written into simultaneously. The I/O graph from the run shows that the redo logs are being read from and written into the archive area. The graphs from the run follow.

Best practices for Oracle 10g and EVA5000
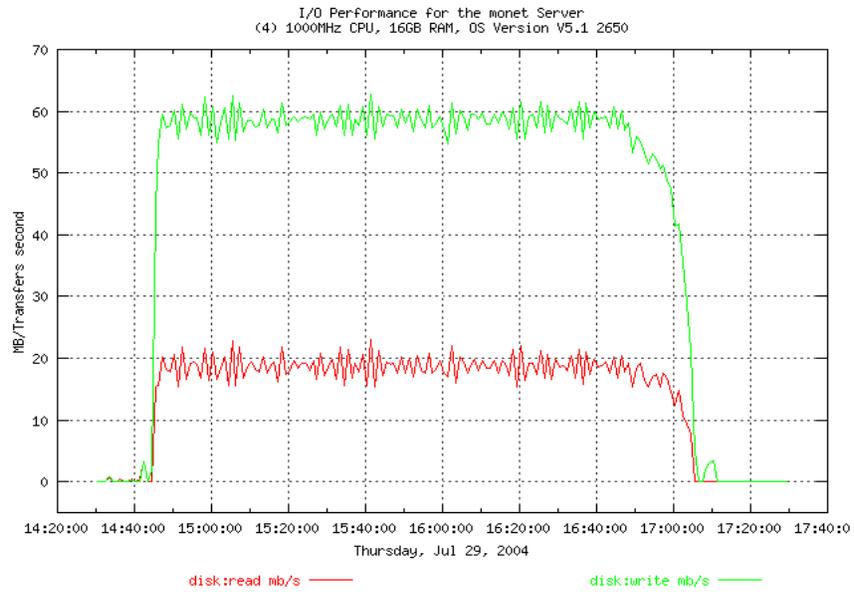
I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Dataload with snapclones – Layout 1

In this test, data was loaded into both databases with archiving turned on and snapclones of the databases were created using SSSU while the load was running. This job was also run twice to get timing from more than one run. Slight variations in the timings for both runs is noticed because the snapclone process was not initiated at the same time during the two runs. They were started a few minutes after the data load began. The following tables list the runtime for these jobs.

Run no. 1

	Start time	End time	Run time
Database A	11:50:52	17:37:01	05:40
Database B	11:50:52	17:37:09	05:40

Run no. 2

	Start time	End time	Run time
Database A	08:27:21	14:31:01	06:04
Database B	08:27:21	14:14:29	05:47

As the CPU utilization and I/O during the tests show, the cloning process has an impact on the workload. The I/O throughput falls considerably and the CPU waits go up dramatically as the EVA uses its resources for creating the snapclones. The workload resumes at its normal levels once the clones are created. The system performance graphs during this run follow.
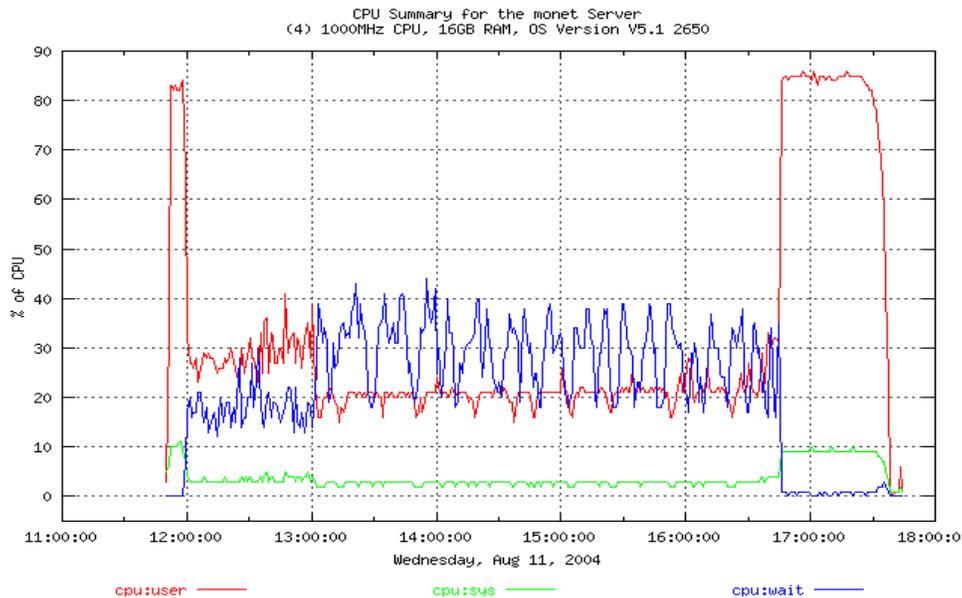
Best practices for Oracle 10g and EVA5000

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650
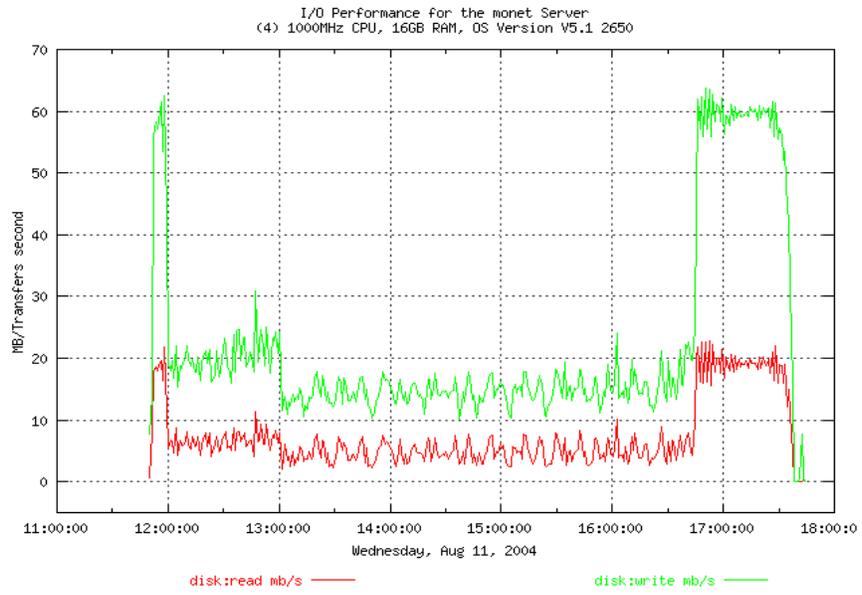
Dataload with snapshots – Layout 1

In this test, data was loaded into both databases with archiving turned on and snapshots of the database units were created using SSSU while the load was running. This job was also run twice to get timing from more than one run. The following tables list the runtime for these jobs.
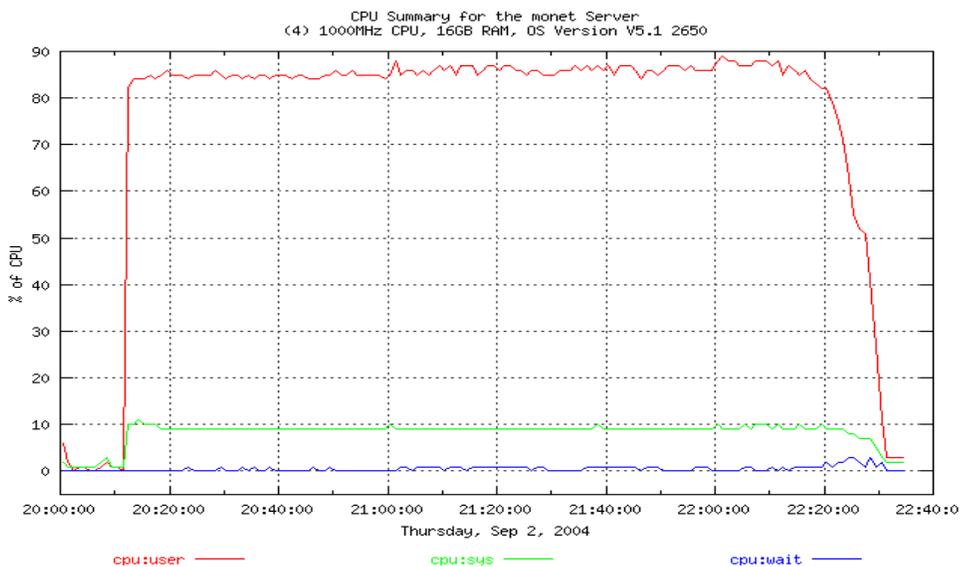
Run no. 1

	Start time	End time	Run time
Database A	20:11:27	22:28:36	02:17
Database B	20:11:29	22:30:09	02:19

Run no. 2

	Start time	End time	Run time
Database A	10:26:02	12:44:32	02:18
Database B	10:25:57	12:47:41	02:22

The snapshot creation process did not have any overheads on the OS. This is clearly seen on the performance graphs and the timing charts, which follow.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ———   disk:write mb/s ———
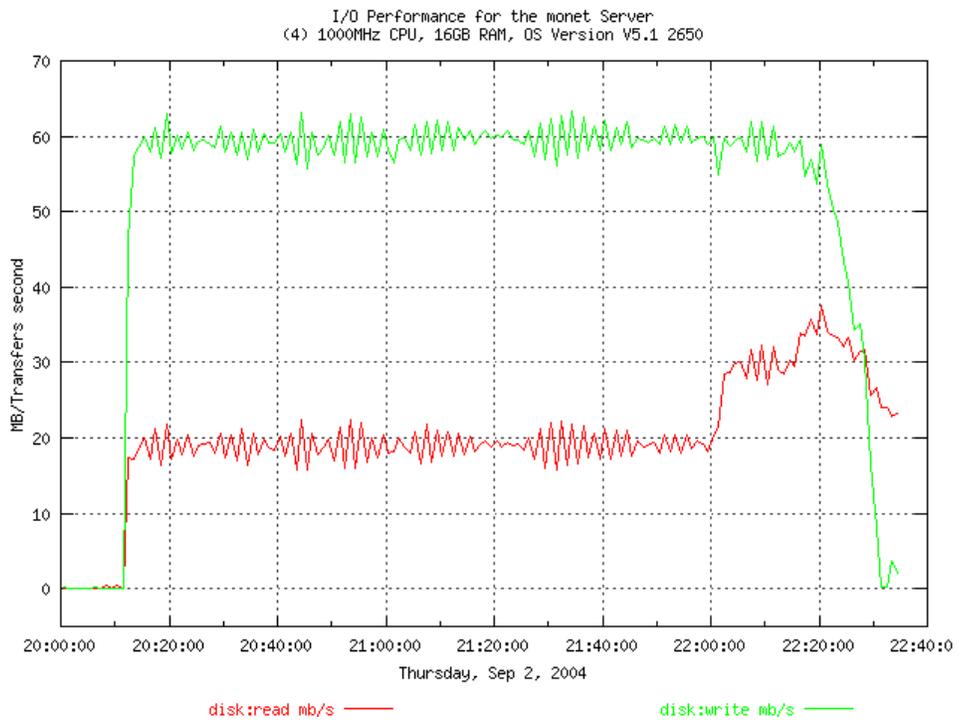
Dataload with restriping – Layout 1

In this test, data was loaded into both the databases with archiving on. Two ungrouped disks were added back into the disk group with the database files. After the restriping was completed, the procedure was repeated again. The following tables list the runtime for these jobs.
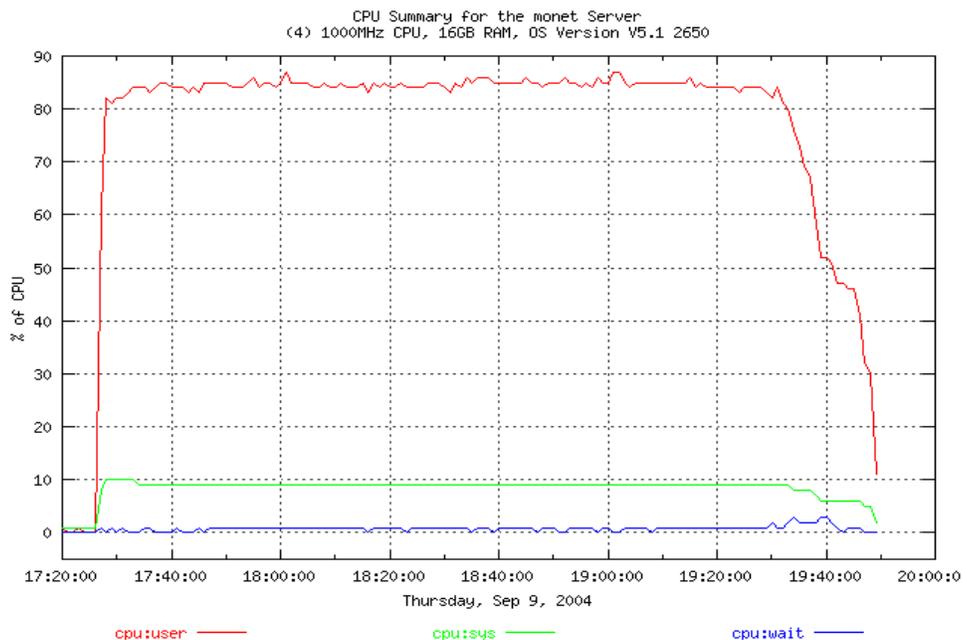
Run no. 1

	Start time	End time	Run time
Database A	17:26:19	19:46:29	02:20
Database B	17:26:22	19:48:56	02:22

Run no. 2

	Start time	End time	Run time
Database A	08:22:38	10:43:47	02:21
Database B	08:22:39	10:42:34	02:20

The restriping process does not affect the OS and the timing of the workload is same as the baseline timing. The system graphs also show no effect of the restriping process. The system performance graphs during this run follow.

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ——   disk:write mb/s ——
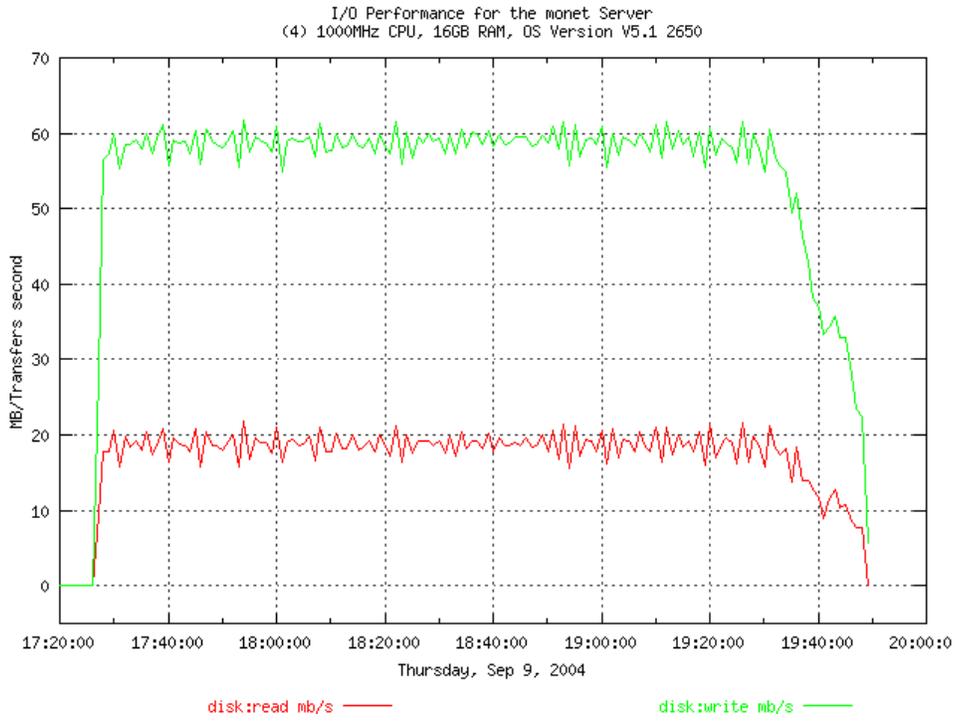
Baseline run without archive logs – Layout 2

In this test, the database workload was run on Layout 2 without archiving. The results from the run were as follows:

	Start time	End time	Run time
Database A	06:28:40	08:53:00	02:25
Database B	06:28:41	08:47:45	02:19

This was similar to the run with Layout 1. The graph output was not collect for this run.

Baseline run with archive logs – Layout 2

In this test, the database workload was repeated twice with archiving turned on. The results were as follows:

Run no. 1

	Start time	End time	Run time
Database A	10:27:55	12:47:58	02:20
Database B	10:27:58	12:48:41	02:21

Run no. 2

	Start time	End time	Run time
Database A	09:42:42	12:01:18	02:19
Database B	09:42:42	12:03:01	02:21

These results were similar to the runs from Layout 1, so the system graphs are not shown in the paper.

Dataload with snapclones – Layout 2

In this test, the workload was repeated with archiving and the database units were cloned while the workload was running. The results from the two runs are shown in the following tables, followed by the system graphs from Run 2. The graphs from Run 1 were not taken as the run time spanned over two different days and the graphs are generated by day.

Run no. 1

	Start time	End time	Run time
Database A	20:50:26	03:36:05	6:46
Database B	20:50:28	03:49:50	6:59

Run no. 2

	Start time	End time	Run time
Database A	15:16:06	22:02:49	6:46
Database B	15:16:07	22:03:00	6:47

As the CPU utilization and I/O during the tests show, the cloning process has an impact on the workload. The I/O throughput falls considerably and the CPU waits go up dramatically as the EVA uses its resources for creating the snapclones. The workload resumes at its normal levels once the clones are created.

Another observation was that with more disk groups, the number of disks in the stripe set is distributed. So the operation takes longer to complete, consequently increasing the run time of the workload.

The system performance graphs during this run follow.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650
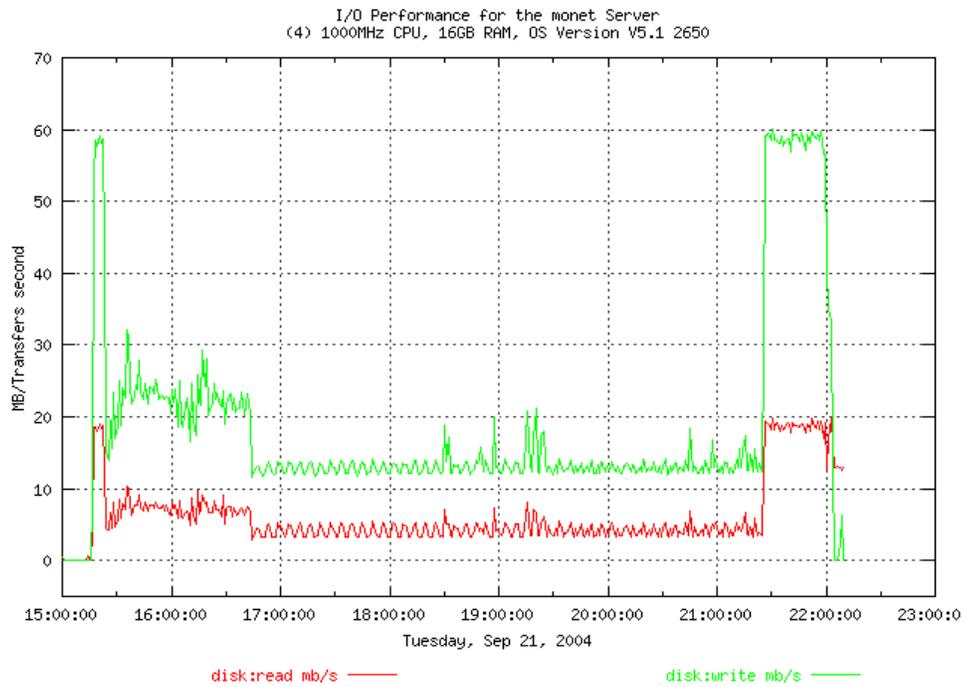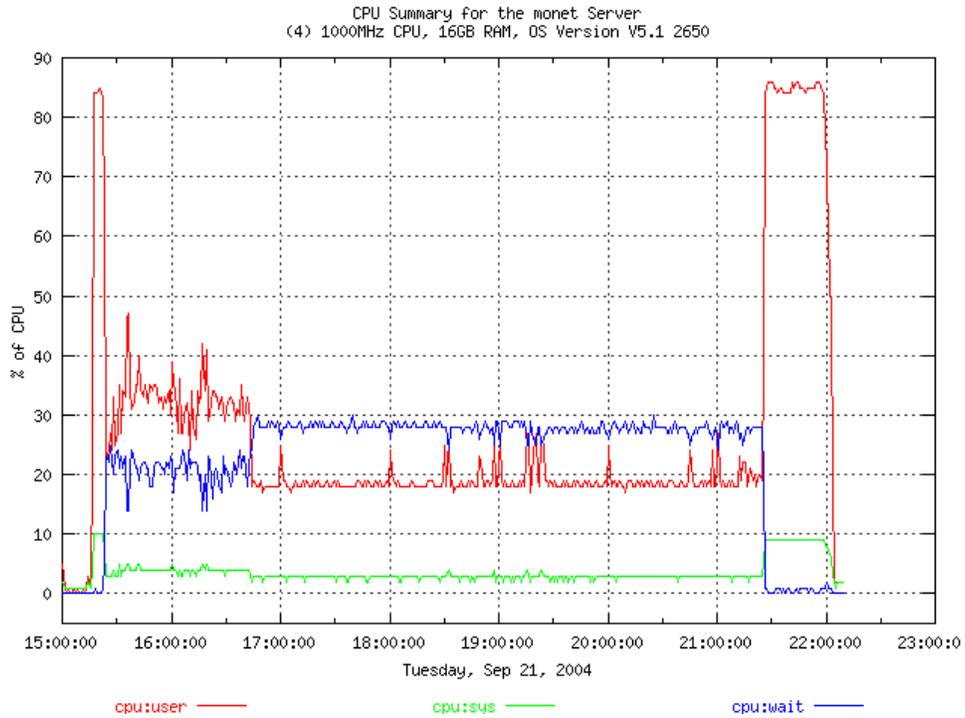
Dataload with snapshots – Layout 2

In this test, the objective of the test was to create the copies of the units in Disk Group 4. However, snapshots have to be created in the same disk group as the source. Since the source disk group did not have enough space, this test was not carried out.

Dataload with restriping – Layout 2

In this test, data was loaded into both the databases with archiving on. Two ungrouped disks were added into the disk group with the database files. After the restriping was completed, the procedure was repeated again. The following tables list the runtime for these jobs.

Run no. 1

	Start time	End time	Run time
Database A	08:23:29	10:41:56	02:18
Database B	08:23:29	10:44:13	02:21

Run no. 2

	Start time	End time	Run time
Database A	20:03:10	22:24:39	02:21
Database B	20:03:09	22:25:52	02:22

The restriping process does not affect the OS and the timing of the workload is same as the baseline timing. The system graphs also show no effect of the restriping process. The system performance graphs during this run follow.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650
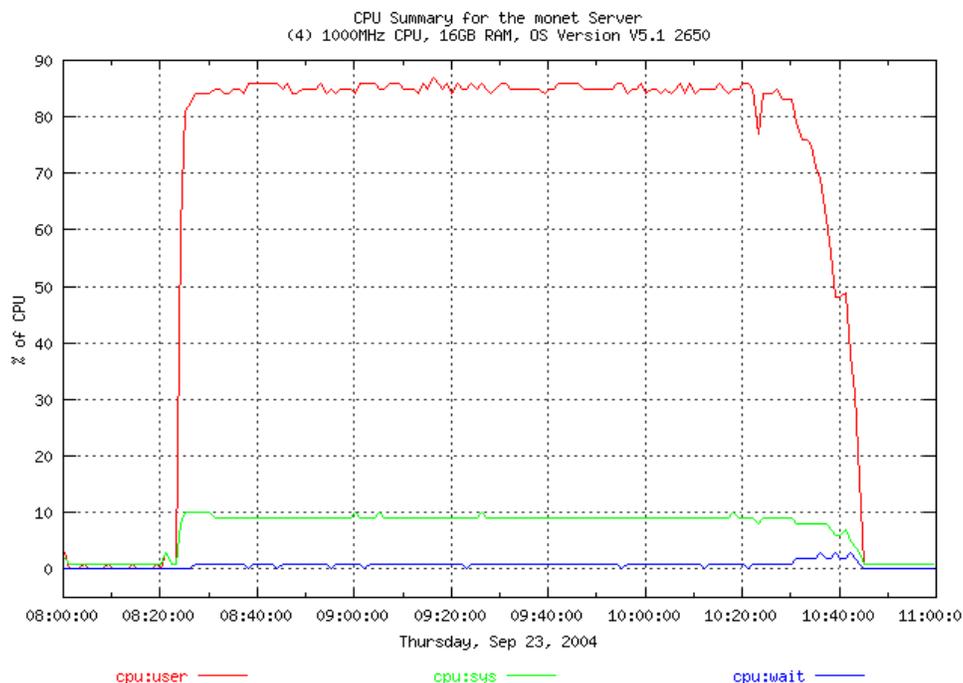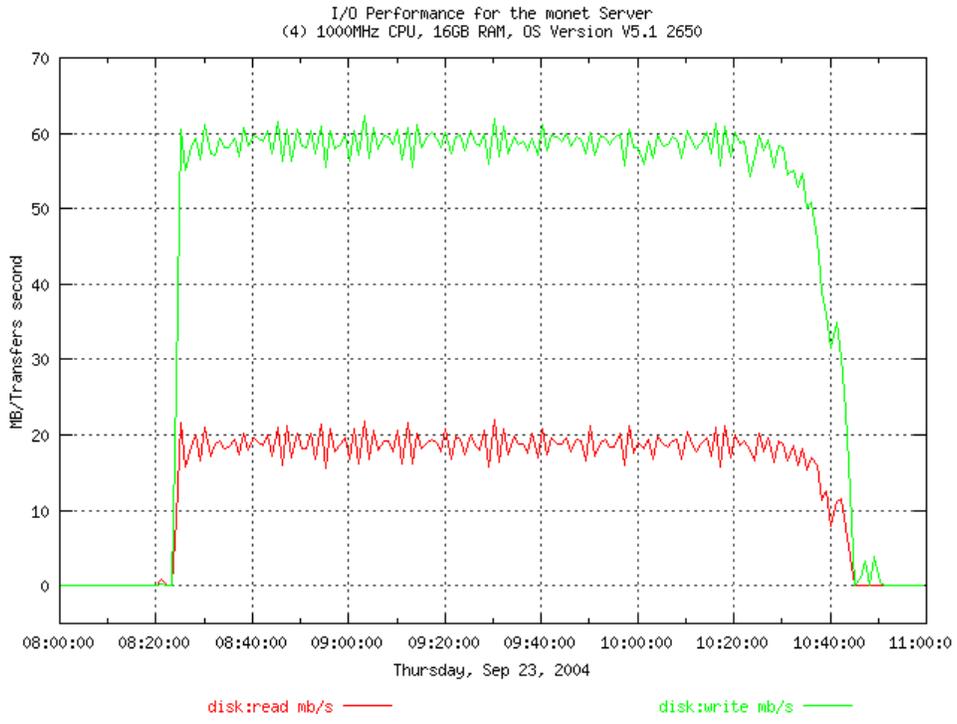
I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Baseline run without archive logs – Layout 3

In this test, the two databases were created in Layout 3. Data was loaded into both databases with archiving turned off. The following tables list the runtime for these jobs.

Run 1

	Start time	End time	Run time
Database A	16:31:26	18:49:26	02:18
Database B	16:31:26	18:50:10	02:19

This was similar to the run with Layout 1. The graph output was not collected for this run.

Baseline run with archive logs – Layout 3

In this test, data was loaded into both databases with archiving turned on. This activity was run twice to verify that runtimes were the same. The runtime for the jobs is as follows

Run 1

	Start time	End time	Run time
Database A	18:56:06	21:17:10	02:21
Database B	18:56:06	21:16:35	02:20

Because this run was also similar to the previous runs, the performance graphs are not shown.

Dataload with snapclones – Layout 3

In this test, data was loaded into both databases with archiving turned on and snapclones of the databases were created using SSSU while the load was running. This job was also run twice to get timing from more than one run. The following tables list the runtime for these jobs.

Run no. 1

	Start time	End time	Run time
Database A	11:58:51	16:59:48	05:01
Database B	11:58:51	17:02:19	05:03

Run no. 2

	Start time	End time	Run time
Database A	06:36:13	11:34:29	04:59
Database B	06:35:12	11:34:10	04:59

CPU utilization and I/O during the tests show that the cloning process has an impact on the workload. The I/O throughput falls considerably and the CPU waits go up dramatically as the EVA uses its resources for creating the snapclones. The workload resumes at its normal levels once the clones are created.

Because there are only two disk groups, the impact of the cloning is quite similar to that of a single disk group.

The system performance graphs during this run follow.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

cpu:user ——   cpu:sys ——   cpu:wait ——

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ——   disk:write mb/s ——
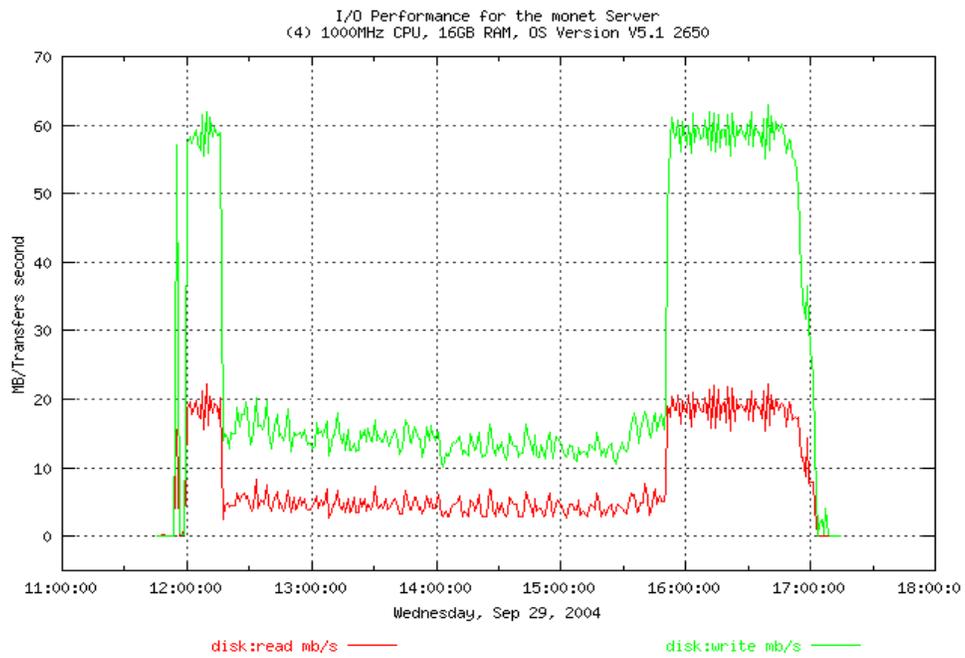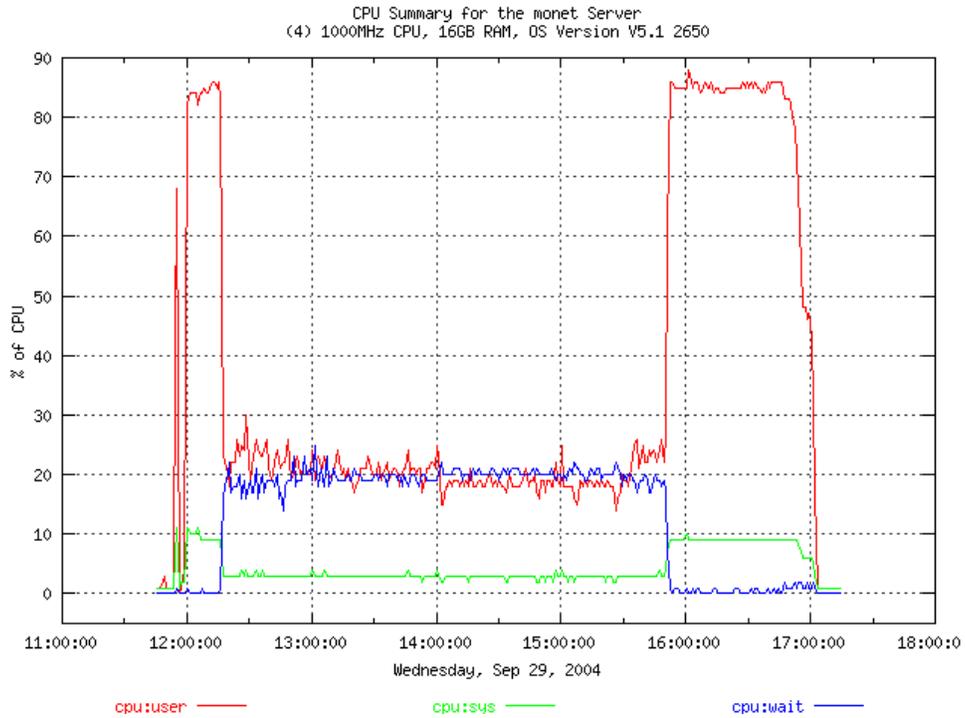
Dataload with
snapshots
– Layout 3

In this test, the objective of the test was to create the copies of the units from one disk group in the other disk group. However, snapshots have to be created in the same disk group as the source. So this test was not carried out.

Dataload with
restriping
– Layout 3

In this test, data was loaded into both the databases with archiving on. Two ungrouped disks were added back into both the disk groups with the database files. After the restriping was completed, the procedure was repeated. The timing from the two runs is captured in the following tables.
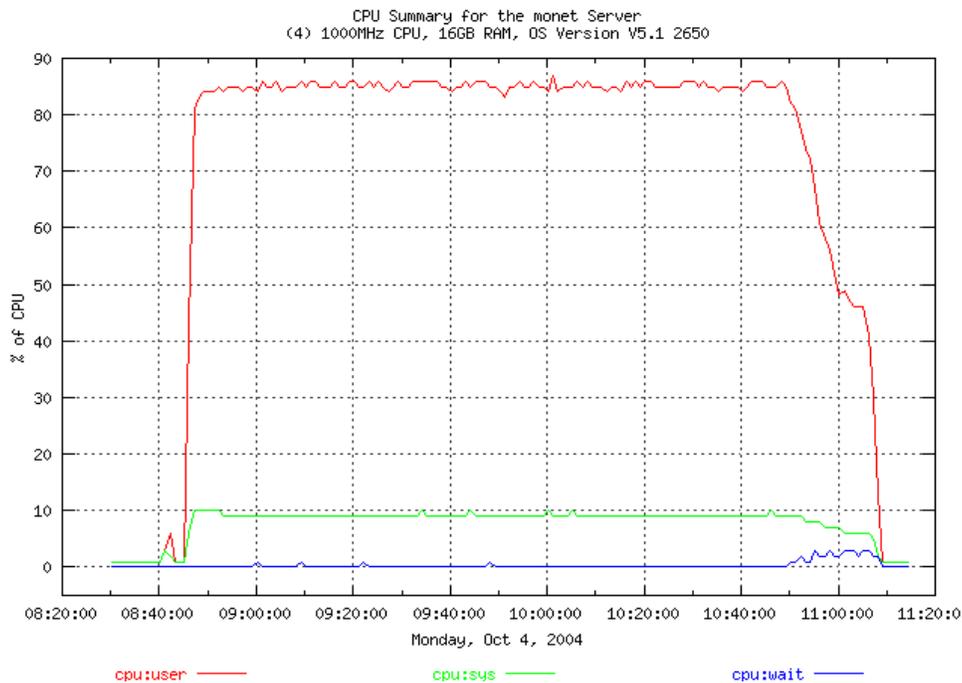
Run no. 1

	Start time	End time	Run time
Database A	08:45:42	11:07:33	02:22
Database B	08:45:43	11:08:19	02:23

Run no. 2

	Start time	End time	Run time
Database A	19:06:19	21:24:28	02:18
Database B	19:06:18	21:27:07	02:21

The restriping process does not affect the OS and the timing of the workload is same as the baseline timing. The system graphs also show no effect of the restriping process. The system performance graphs during this run follow.

24

Best practices for Oracle 10g and EVA5000

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Monday, Oct 4, 2004

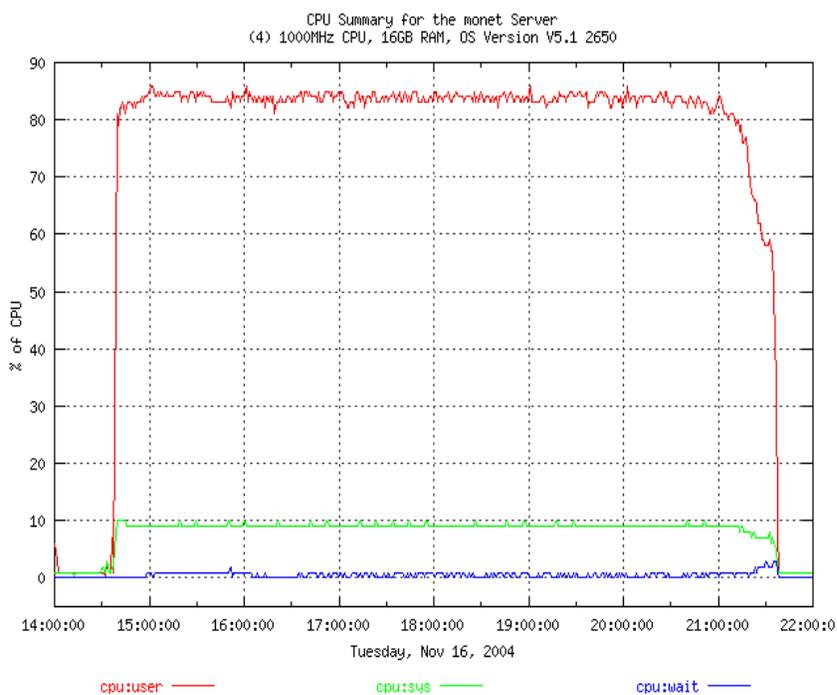disk:read mb/s ———   disk:write mb/s ———
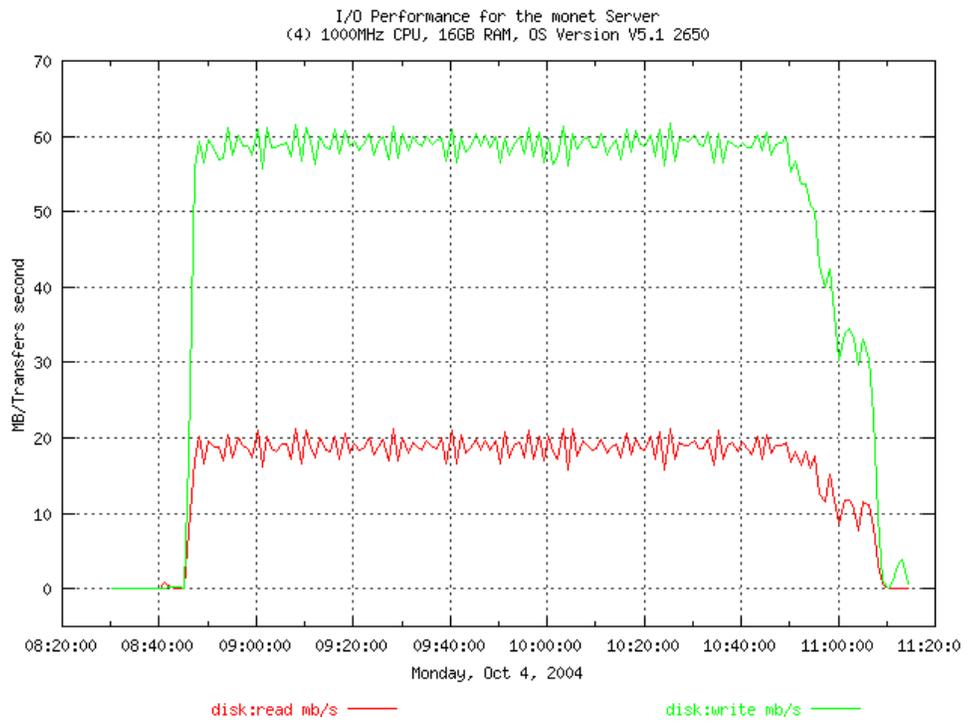
Phase II testing

In Phase 2, the EVAs were running VCS 3.0.2.0. The cloning and snapshotting operations were completed using Business Copy software.
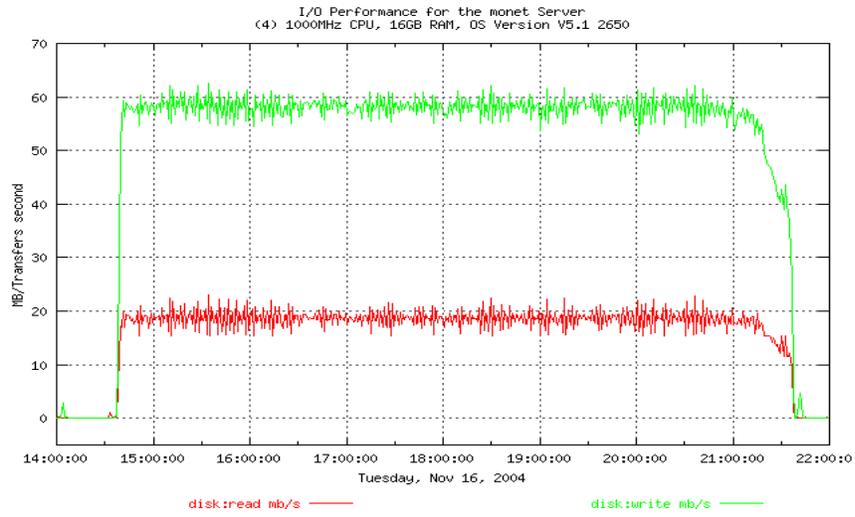
Baseline run with archive logs – Layout 1

In this test, data was loaded into both databases with archiving turned on. This activity was run twice to verify that runtimes were the same. The following tables list the runtime for these jobs.

	Start time	End time	Run time
Database A	14:38:02	21:36:54	06:58
Database B	14:38:03	21:36:38	06:58

By ensuring that the archive logs are on separate disks and by distributing the redo logs, we ensure that no units are being read from or written into simultaneously. As can be seen in the following I/O graph, the redo logs are being read from and written into the archive area.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Best practices for Oracle 10g and EVA5000

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ———   disk:write mb/s ———

Dataload with snapclones – Layout 1

In this test, data was loaded into both databases with archiving turned on and snapclones of the databases were created using a BC job while the load was running. This job was also run twice to get timing from more than one run. As soon as the BC job was completed, the snapclones were mounted on another host and the data was backed to null devices using dd. The following tables list the runtime for these jobs.
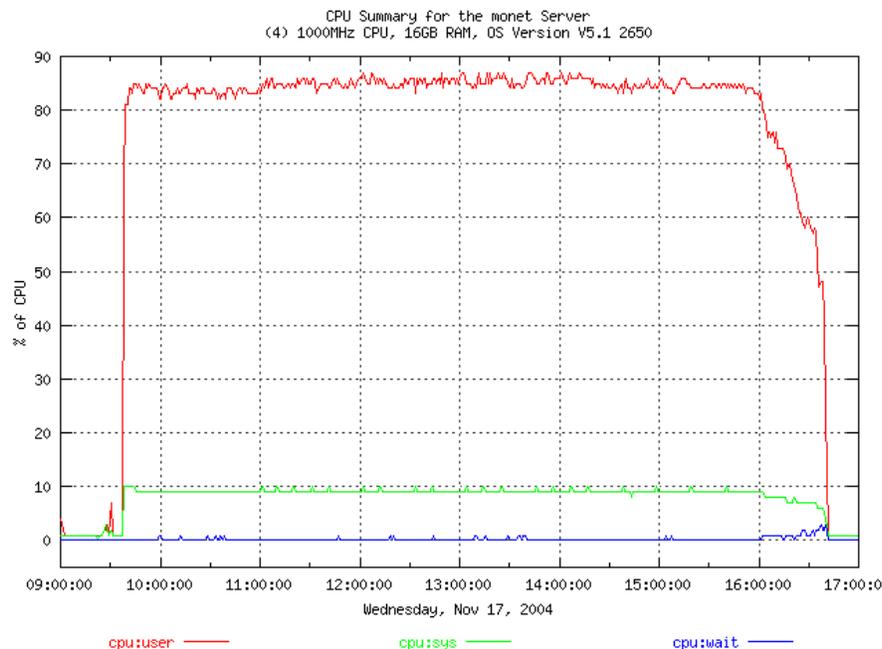
Run no. 1

	Start time	End time	Run time
Database A	09:37:25	16:40:22	07:03
Database B	09:37:25	16:39:32	07:02

Run no. 2

	Start time	End time	Run time
Database A	10:00:59	16:49:48	06:49
Database B	10:01:00	16:53:24	06:52

Unlike the snapclone process performed with VCS 3.0.1.0, the snapclone process with VCS 3.0.2.0 did not put any overhead on the system performance. The performance graphs show the same system behavior as the system running without any EVA operations. The system performance graphs during this run follow.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Wednesday, Nov 17, 2004

cpu:user ——   cpu:sys ——   cpu:wait ——

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ——————   disk:write mb/s ——————
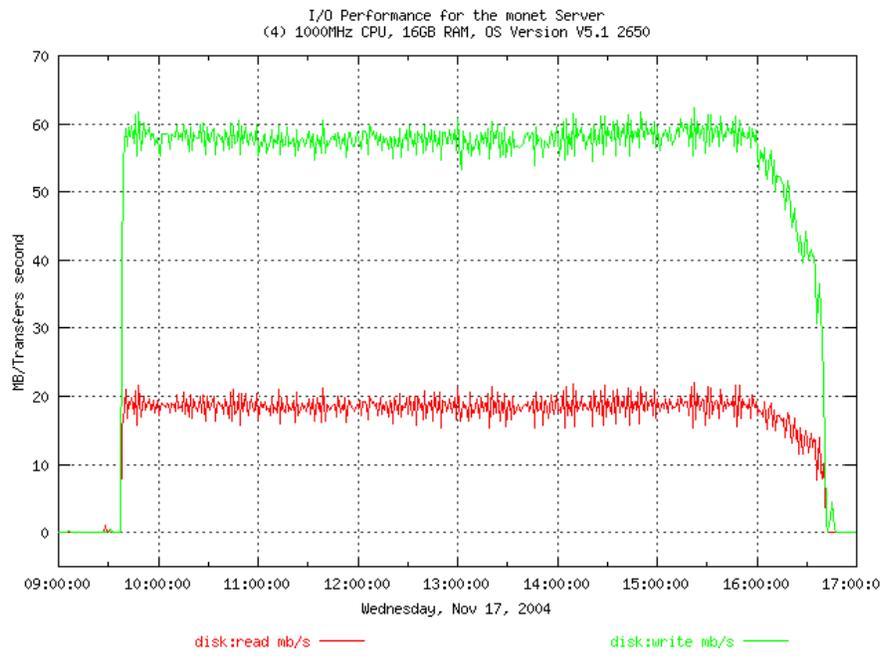
Dataload with On-demand snapshots – Layout 1

In this test, data was loaded into both databases with archiving turned on and on-demand snapshots of the database units were created with BC while the load was running. As soon as the BC job was completed, the snapshots were mounted on another host and the data was backed to null devices using dd. This job was also run twice to get timing from more than one run. The following tables list the runtime for these jobs.
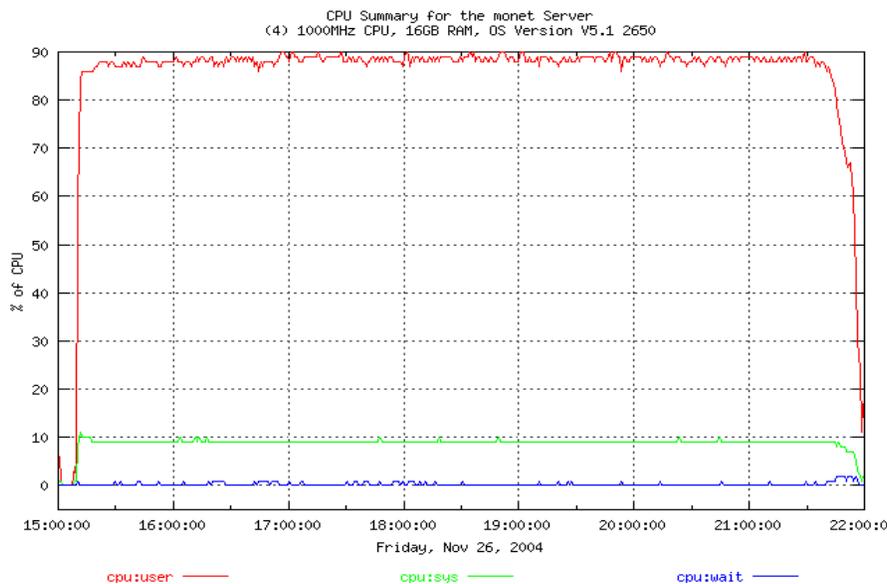
Run no. 1

	Start time	End time	Run time
Database A	19:48:21	02:46:02	06:58
Database B	19:48:19	02:45:15	06:57

Run no. 2

	Start time	End time	Run time
Database A	15:09:39	21:55:48	06:46
Database B	15:09:41	21:56:09	06:47

The snapshot creation process and taking backup from the snapshots from another host had no overhead on the OS running the Oracle job. The system performance graphs from Run 2 follow.
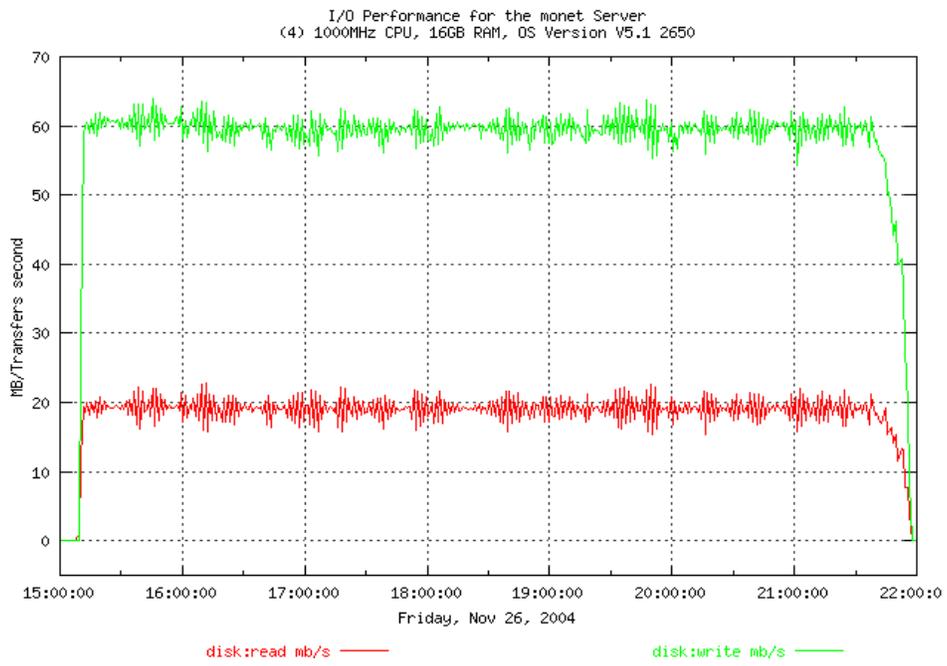
I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

MB/Transfers second

| | 15:00:00 | 16:00:00 | 17:00:00 | 18:00:00 | 19:00:00 | 20:00:00 | 21:00:00 | 22:00:0 |

Friday, Nov 26, 2004

disk:read mb/s ———   disk:write mb/s ———

Dataload with fully allocated snapshots – Layout 1

In this test, data was loaded into both databases with archiving turned on and fully allocated snapshot of the database units were created with BC while the load was running. As soon as the BC job was completed, the snapshots were mounted on another host and the data was backed to null devices using dd.   This job was also run twice to get timing from more than one run. The following tables list the runtime for these jobs.
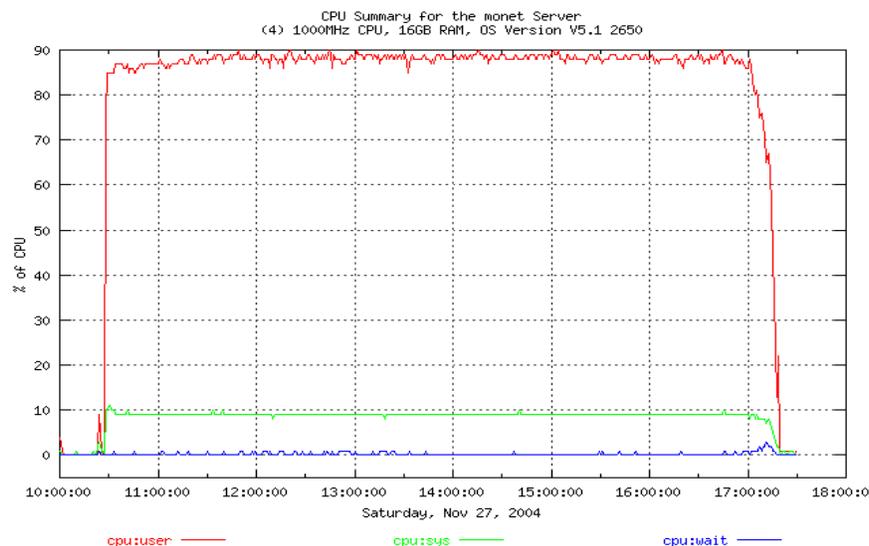
Run no. 1

	Start time	End time	Run time
Database A	10:27:22	17:14:08	06:47
Database B	10:27:24	17:16:40	06:49

Run no. 2

	Start time	End time	Run time
Database A	06:55:13	13:46:38	06:51
Database B	06:55:16	13:44:41	06:49

The snapshot creation process and taking backup from the snapshots from another host had no overhead on the OS running the Oracle job. The system performance graphs during this run follow.

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650
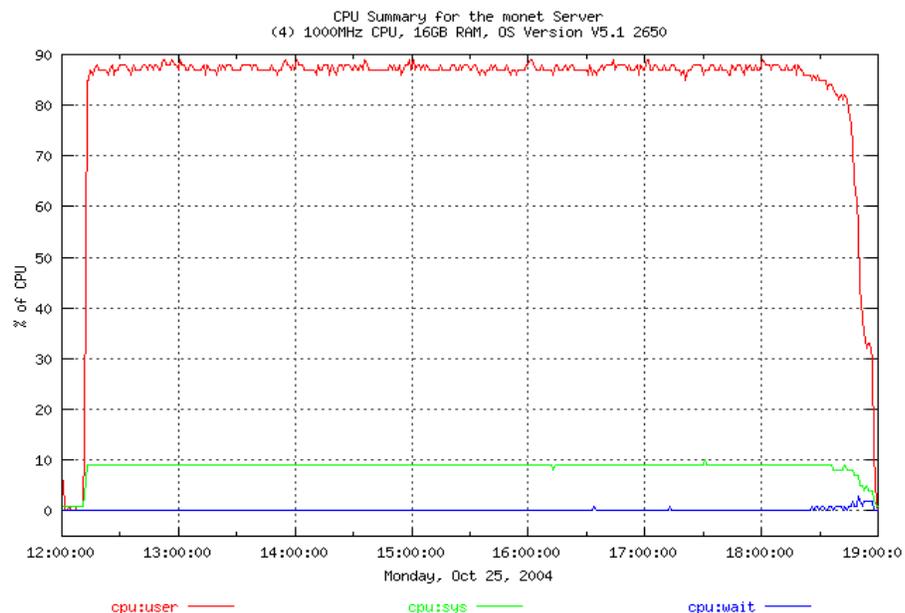
disk:read mb/s ——   disk:write mb/s ——
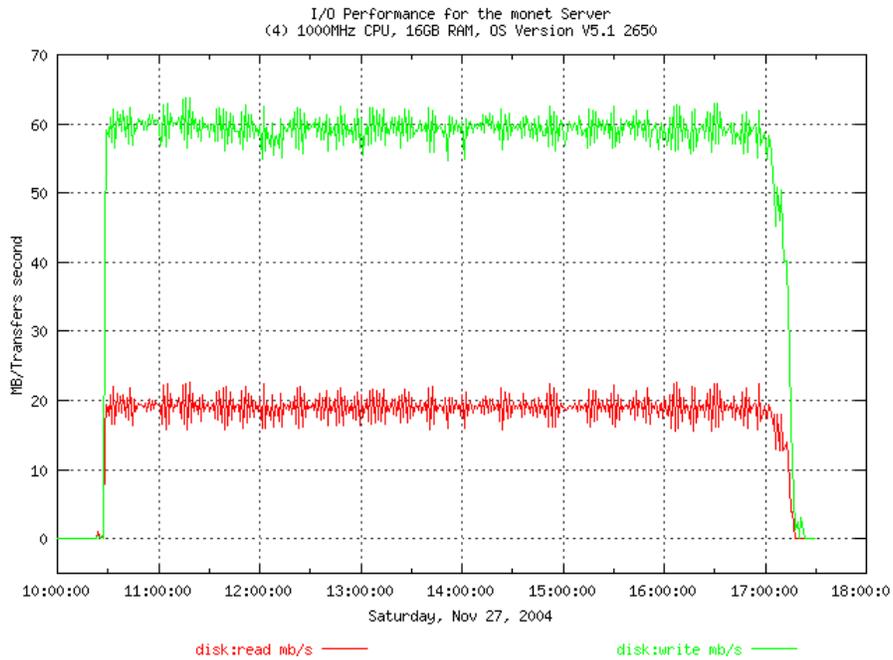
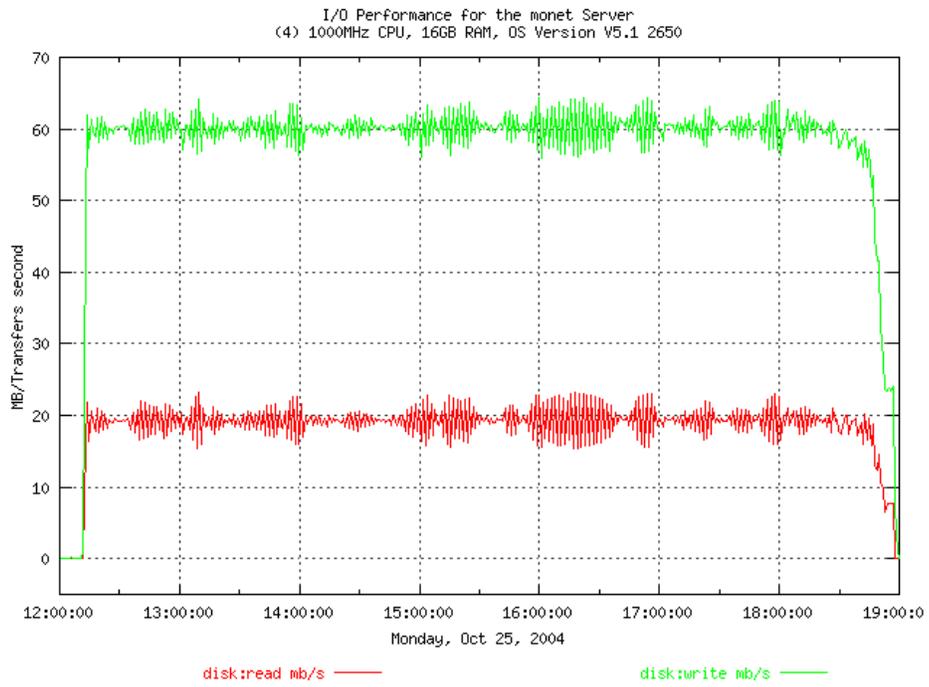**Baseline run with archive logs – Layout 3**

In this test, data was loaded into both databases with archiving turned on. This activity was run twice to verify that runtimes were same. The following tables list the runtime for these jobs.

	Start time	End time	Run time
Database A	12:11:46	18:53:13	06:42
Database B	12:11:45	18:57:40	06:46

By ensuring that the archive logs are on separate disks and by distributing the redo logs, we ensure that no units are being read from or written into simultaneously. As can be seen in the I/O graph, the redo logs are being read from and written into the archive area. The graphs from the run follow.

CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

cpu:user ——   cpu:sys ——   cpu:wait ——

33

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ——   disk:write mb/s ——

Monday, Oct 25, 2004

Dataload with snapclones -- Layout 3

In this test, data was loaded into both databases with archiving turned on and snapclones of the databases were created using a BC job while the load was running. This job was also run twice to get timing from more than 1 run. As soon as the BC job was completed, the snapclones were mounted on another host and the data was backed to null devices using dd. The following tables list the runtime for these jobs.
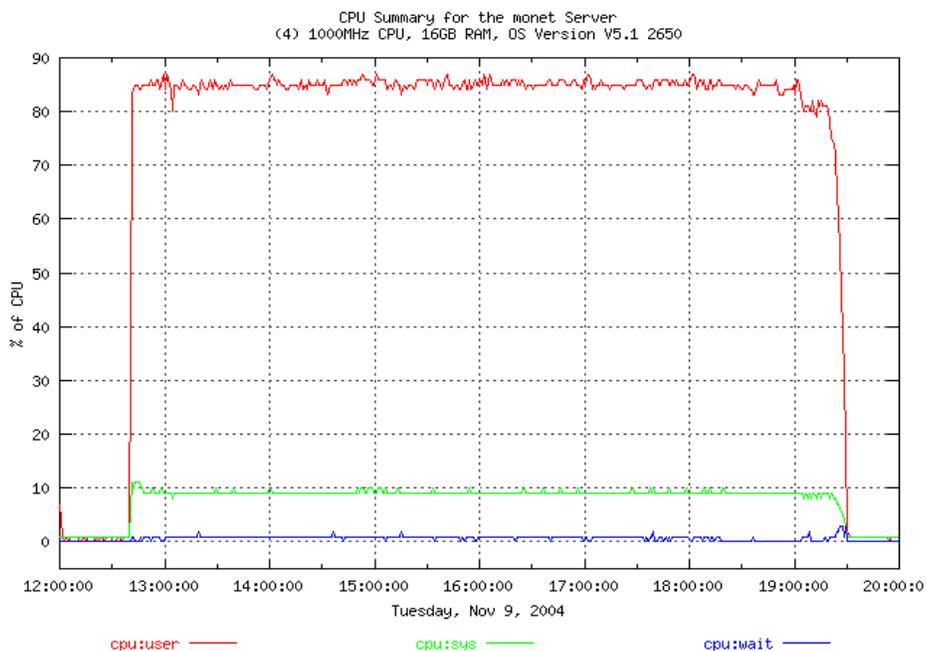
Run no. 1

	Start time	End time	Run time
Database A	12:39:53	19:29:36	06:50
Database B	12:39:53	19:28:24	06:49

Run no. 2

	Start time	End time	Run time
Database A	07:11:39	14:08:44	06:57
Database B	07:11:38	14:05:01	06:54

The system performance graphs during this run follow:

Best practices for Oracle 10g and EVA5000

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ———   disk:write mb/s ———
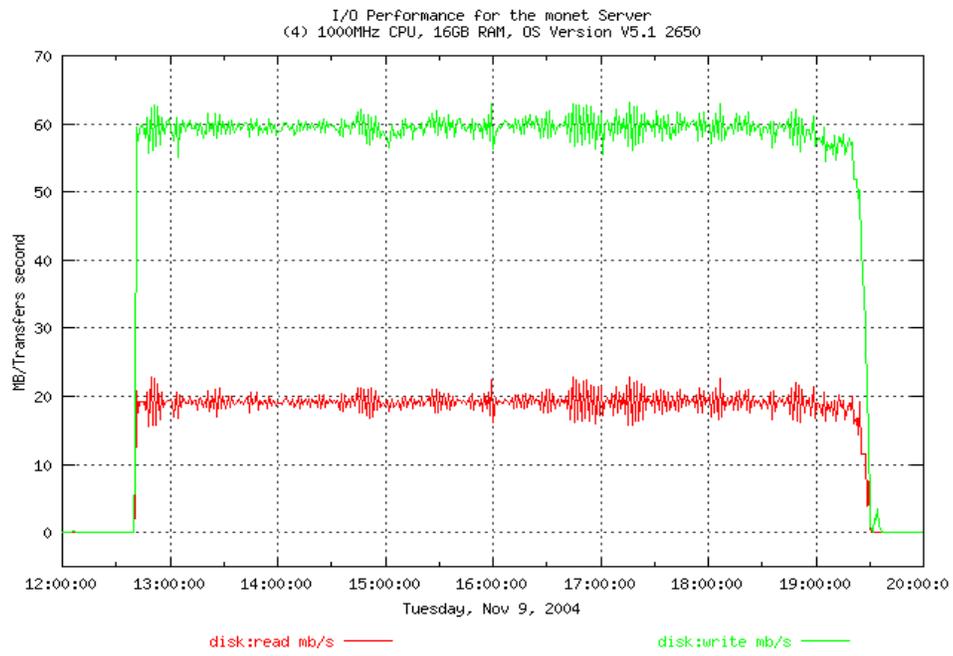
Dataload with On-demand snapshots – Layout 3

In this test, data was loaded into both databases with archiving turned on and on-demand snapshot of the database units were created with BC while the load was running. As soon as the BC job was completed, the snapshots were mounted on another host and the data was backed to null devices using dd.   This job was also run twice to get timing from more than one run. The following tables list the runtime for these jobs.
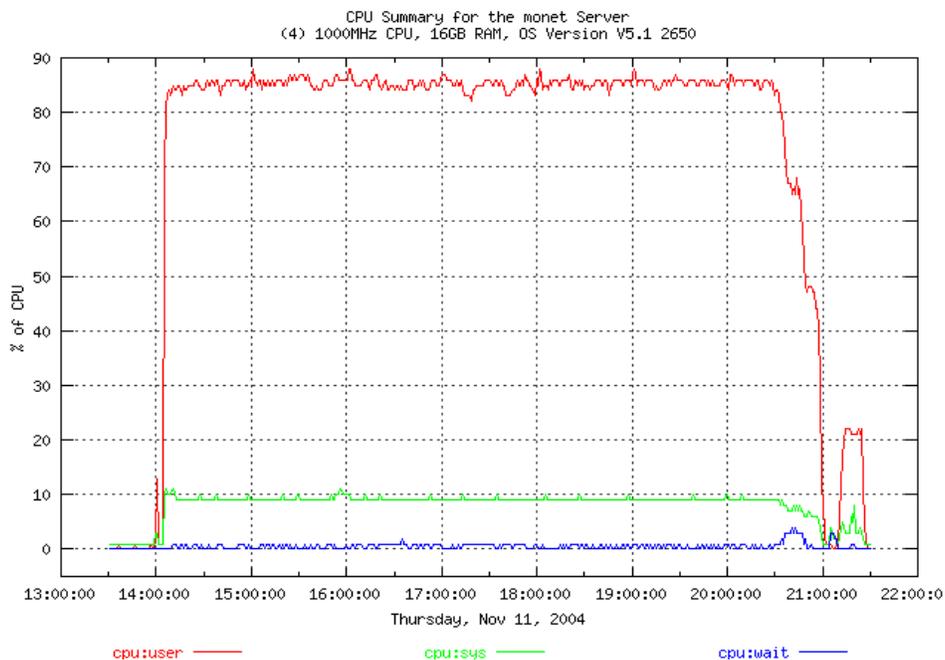
Run no. 1

	Start time	End time	Run time
Database A	14:04:29	20:57:44	06:53
Database B	14:04:29	20:59:27	06:55

Run no. 2

	Start time	End time	Run time
Database A	07:19:07	14:04:14	06:45
Database B	07:19:07	14:00:08	06:41

The snapshot creation process and taking backup from the snapshots from another host have no overhead on the OS running the Oracle job. The system performance graphs during this run are follow.
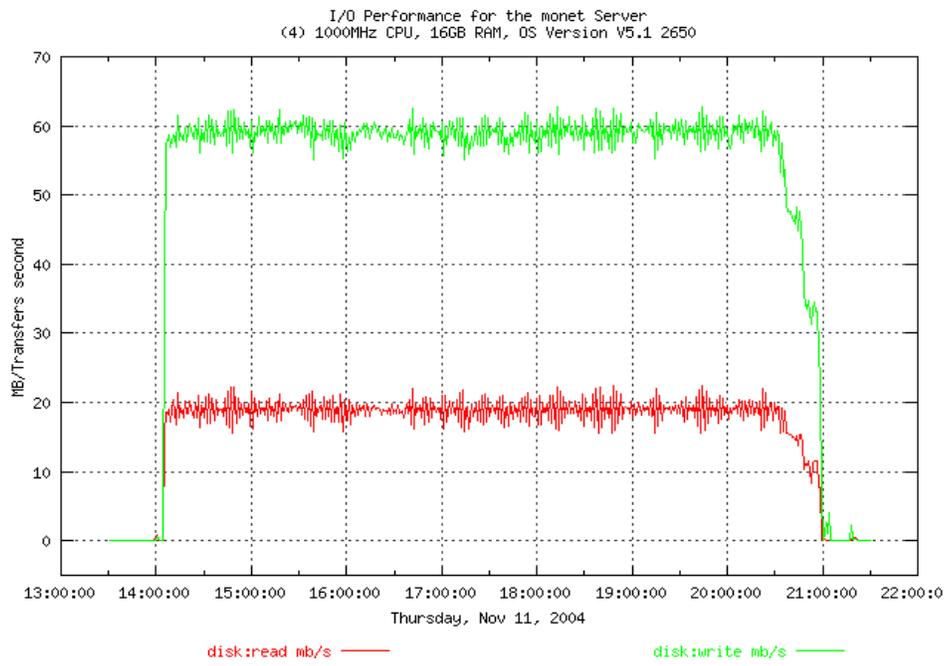
CPU Summary for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

37

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

Thursday, Nov 11, 2004

disk:read mb/s ———   disk:write mb/s ———

Dataload with fully allocated snapshots – Layout 3

In this test, data was loaded into both databases with archiving turned on and fully allocated snapshot of the database units were created with BC while the load was running. As soon as the BC job was completed, the snapshots were mounted on another host and the data was backed to null devices using dd. This job was also run twice to get timing from more than one run. The following tables list the runtime for these jobs.
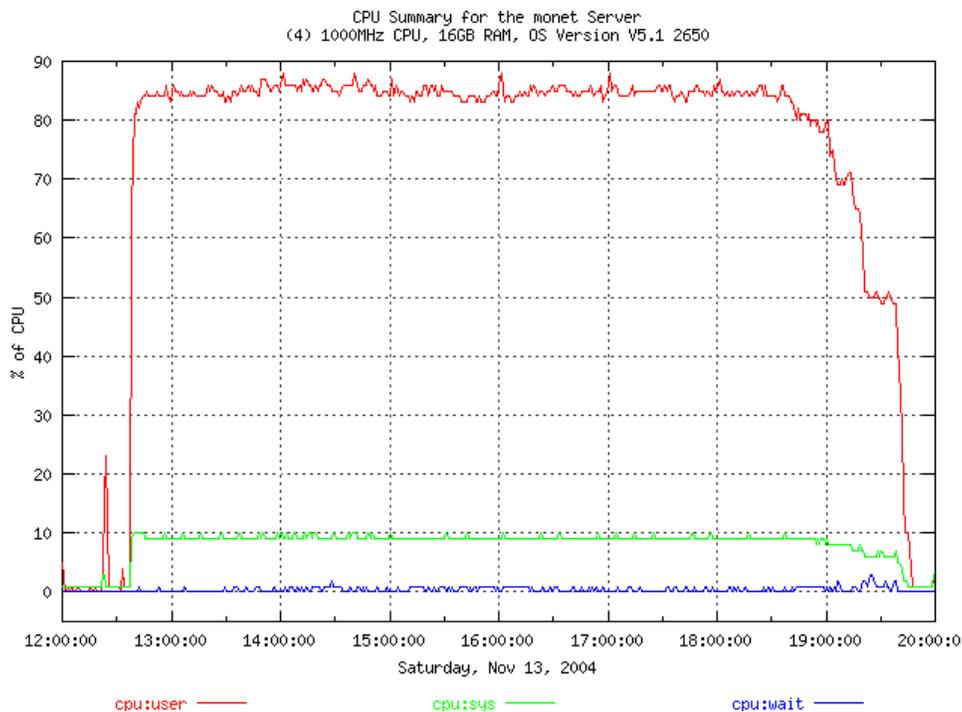
Run no. 1

	Start time	End time	Run time
Database A	12:37:32	19:45:54	07:08
Database B	12:37:32	19:38:25	07:01

Run no. 2

	Start time	End time	Run time
Database A	09:23:56	16:21:42	06:58
Database B	09:23:57	16:20:41	06:57

The snapshot creation process and taking backup from the snapshots from another host had no overhead on the OS running the Oracle job. The system performance graphs during this run follow:

I/O Performance for the monet Server
(4) 1000MHz CPU, 16GB RAM, OS Version V5.1 2650

disk:read mb/s ——   disk:write mb/s ——
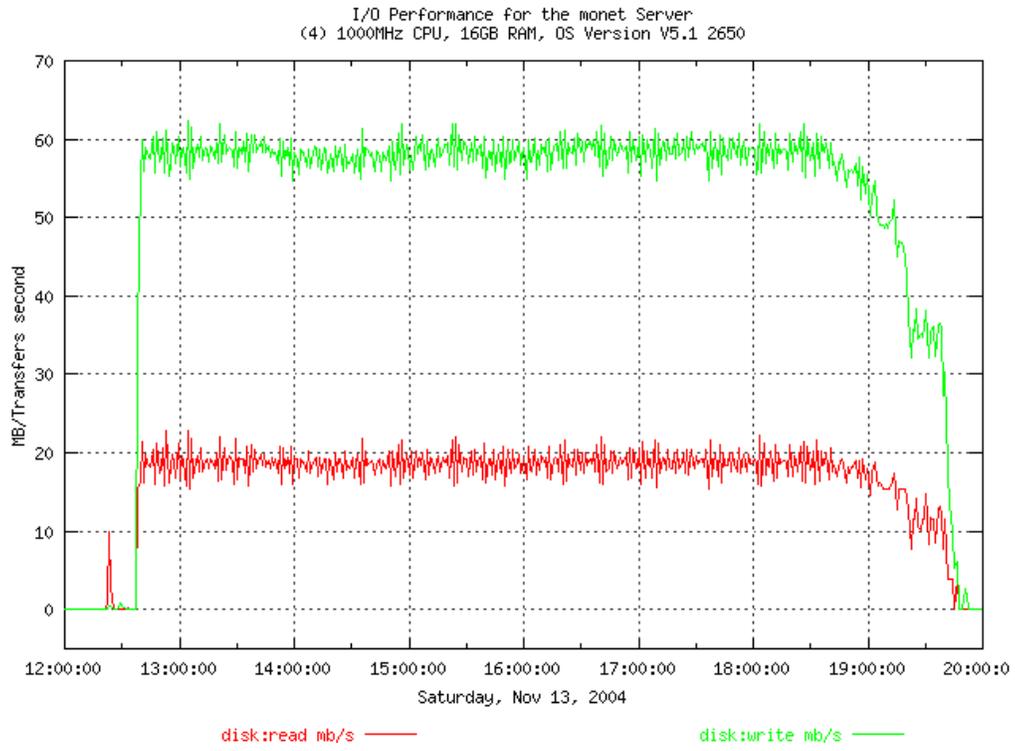
Summary

The different scenarios tested help us arrive at the following conclusions.

- Separate the Oracle data files, redo log files, rollback (undo) files and archive log files. Putting each redo log files on it's own file system helps as each file is reread during the archival process. This helps avoid simultaneous reads and writes on the file systems with the redo logs as the redo log behavior is predictable unlike the data files.

- Use RAID 1 virtual disks should be used for redo logs as they deliver the best performance for sequential writes.

- Database files can be stored on VRAID 5 disks as read performance on RAID 5 is on par with read performance on VRAID 1. The writes to data files are done asynchronously by the Oracle instance and VRAID 5 does not cause bottleneck for database writes. The only time VRAID 5 can run slower than VRAID 1 is during file creates which is few and far between.

- For most Oracle jobs, not much difference was seen in workload performance with different disk group layouts. However, there was a noticeable difference in performance for EVA functions. This was noticed when more disk groups were used. This is because of the way the EVA stripes data across all disks in each disk group. Having a single disk group allows for striping across more spindles thereby reducing the possibility of data access conflicts. For high availability and avoiding the risk of losing database files and redo logs at the same time due to disk group failure, separating them is a best practice. Put the redo logs and archive logs under 1 disk group and all the files can go on a single group. If more than 1 EVA is being used, and other EVAs are used for datafiles only, then a single disk group is recommended.

- Snapclones created on EVAs running VCS 3.0.1.0. affected system performance. It caused I/O waits while the snapclone process was running. However, no effect was noticed during snapshot creation.

- There is no overhead for creating snapshots (on demand or fully allocated) or snapclones, mounting on another host and backing data from the clones / snapshots when the EVAs were running VCS 3.0.20.

- There was no overhead for Restriping operations when new disks were added to existing disk groups. This is especially useful to remove disks and adding in larger disks into disk groups to increase storage capacity.

- For large tablespaces, use locally managed tablespaces wherever possible as block management within the tablespace is done without any changes to the Oracle data dictionary, thereby reducing rollback transactions. This can be beneficial for databases doing lots of inserts / deletes. The difference between locally managed tablespaces vs dictionary managed tablespaces can be found in the Oracle administrator's guide

- Snapshots and Snapclones are created in the same disk group as the original disk group using BC software.

- If file systems are used for Oracle, follow the guidelines defined in the OFA (Oracle Flexible Architecture). The next section describes OFA and lists some of its advantages.

Oracle Flexible Architecture

The OFA standard is a set of configuration guidelines for creating fast, highly available, reliable Oracle databases that require little maintenance. Some of the characteristics of an OFA compliant database are:

- Organizes file systems to allow for easy administration
- Accommodates scalability, such as adding data into existing databases, or adding users to the system
- Distributes I/O loads across disk drives to prevent performance bottlenecks caused by multiple read/write commands issued simultaneously to a single drive
- Distributes applications across more than one drive to safeguard against disk failure
- Ensures integrity of login home directories when home directories are added, moved, or deleted by the DBA
- Allows multiple versions of the application software for concurrent execution of application software tasks

Business Copy

Business copy (BC) is an application that allows users to make point in time copies of storage units on EVA,MA and EMA. The copies known as Business Continuance Volumes (BCVs) can be mounted on any supported operating system in the BC network.

For more information on BC, refer to "hp StorageWorks Business Copy eva/ma/ema."

SSSU

The Storage System Scripting Utility (SSSU) for Enterprise Virtual Array is a command line application that allows you to configure and control EVA 3000 and 5000 controllers. You can execute configuration requests using Command View EVA or SSSU.

Command View EVA is the graphical user interface that allows you to control and monitor a storage system. Use Command View EVA to handle simple or initial configuration tasks easily and expediently. Use SSSU to script and run repetitious and complex configuration tasks.

For more information, refer to the "hp StorageWorks Storage System Scripting Utility Reference Guide."