

# Memory file system (MemFS) for HP-UX 11i v3



**Installation, configuration and tuning**

HP Integrity – The Most Trusted. Always.



©2008 Hewlett-Packard Development Company, L.P.  
The information contained herein is subject to change without notice.

# Introducing the speaker

- Alban Lyndem is project lead for the *Memory File System* implementation for HP-UX 11i v3.
- Alban works in the HP Operating Systems Lab.



# Contents

- Overview of MemFS
  - Memory based file system and its usage
  - Design overview and limitations of HP-UX 11i v2 MemFS
  - Design overview and advantages of HP-UX 11i v3 MemFS
  - Comparison of HP-UX 11i v3 MemFS with other memory file system products ( e.g. HP-UX 11i v2 MemFS, Solaris tmpfs, Linux tmpfs, VxFS )
  - Features available
  - Performance

# Contents (continued)

- Installation, configuration, and tuning of HP-UX 11i v3 MemFS (contents)
  - How to Install/Configure HP-UX 11i v3 MemFS
  - How to Create and Use HP-UX 11i v3 MemFS
  - Tuning the HP-UX 11i v3 MemFS
  - Guidelines
  - Unsupported features
  - Documentation

## MemFS overview

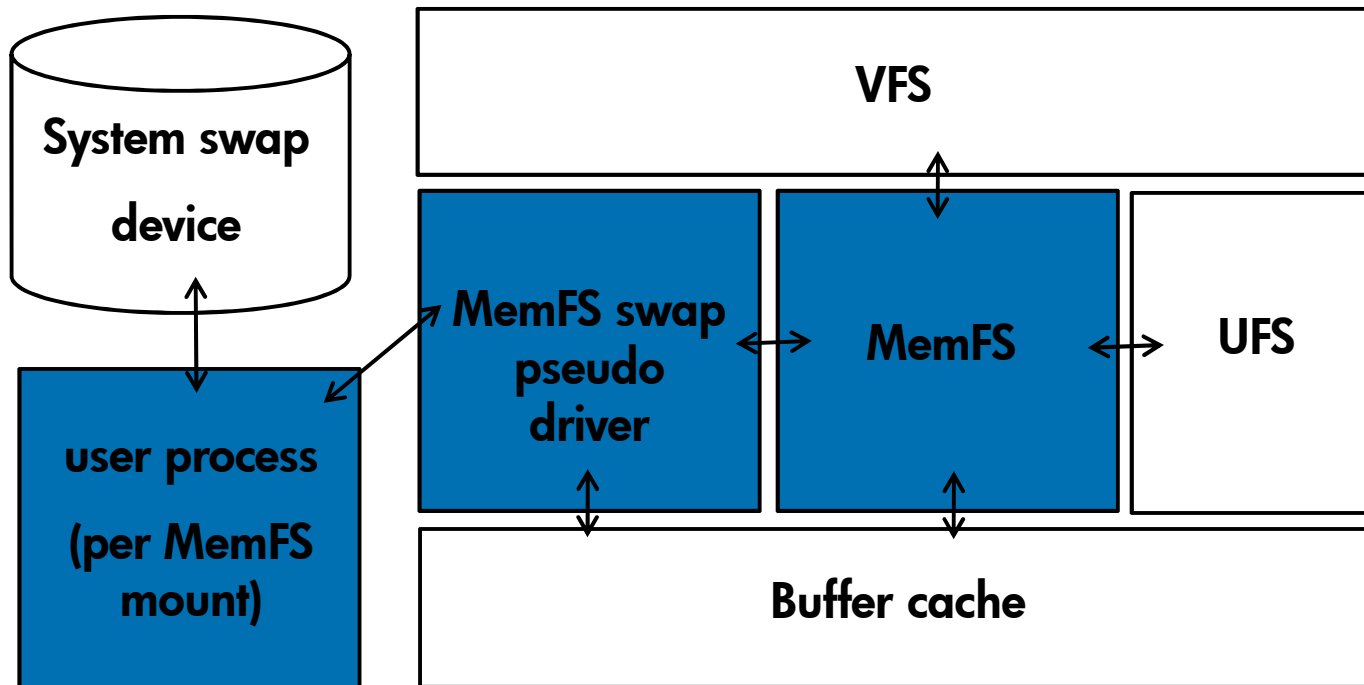


# Memory-based file system and its usage

- A file system in system physical memory
- Its metadata and data are not backed up by any persistent storage device such as disks
- Under system memory pressure, its data is swapped into the system swap device
- Advantages
  - Storage for temporary files
  - For getting better performance with the applications which does extensive metadata operations like creation, deletion of files and directories
- Disadvantages
  - Can not be used as a replacement for disk based file system as it does not preserve data across mounts/reboots
  - Excessive use of system memory by it can adversely affect other memory consumers



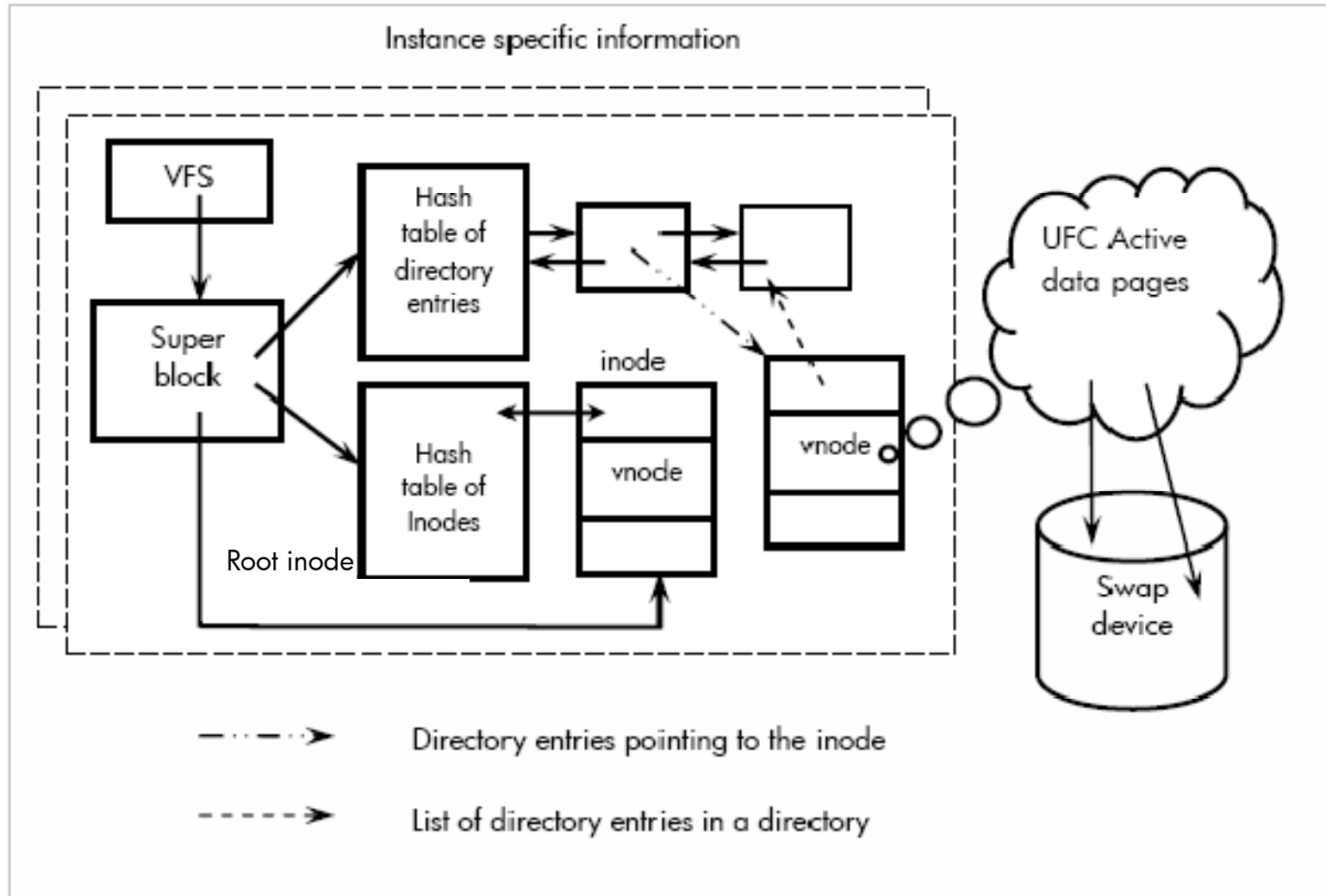
# Design overview of HP-UX 11i v2 MemFS



# Limitations of HP-UX 11i v2 MemFS

- File size and File system size are restricted by UFS file system limits (256 GB)
- No. of files that can be created on a MemFS instance is restricted by file system size
- All the inodes are pre-allocated, wasting kernel memory
- Scalability issues with large directories
- Overheads during allocation of blocks for files and directories
- Pre-allocates virtual memory of file system size (similar to swappable RAM disk ) during the creation of file system, which can cause the system to run out of virtual memory
- Swapping in/out of MemFS data is very slow

# Design overview of HP-UX 11i v3 MemFS



# Competitive data

Feature	HP-UX 11i v3 MemFS	HP-UX 11i v2 MemFS	Tmpfs (Solaris & Linux)	MFS (Tru64 UNIX)	HP-UX RAMdisk
Architecture	Unified file cache (UFC only)	Buffer Cache (11.23) and user process address space	Kernel anonymous memory	User process address space of newfs process	User process address space
Swap specification	Implicitly uses VM swap. Cannot be changed	Implicitly uses VM swap. Cannot be changed	Implicitly uses VM swap. Cannot be changed.	Implicitly uses VM swap. Cannot be changed	Implicitly uses VM swap. Cannot be changed
Configuration	One step. Mount (manual) or /etc/fstab (on boot)	One step. Mount (manual) or /etc/fstab (on boot)	One step. Mount (manual) or /etc/fstab (on boot)	Two steps. newfs and then mount.	Three steps. RAMdisk creation, newfs and then mount
Maximum Memory Footprint	Entire Memory available for UFC	Limited to process address space	Entire Physical Memory or tunable specified	Limited to process address space	Limited to process address space

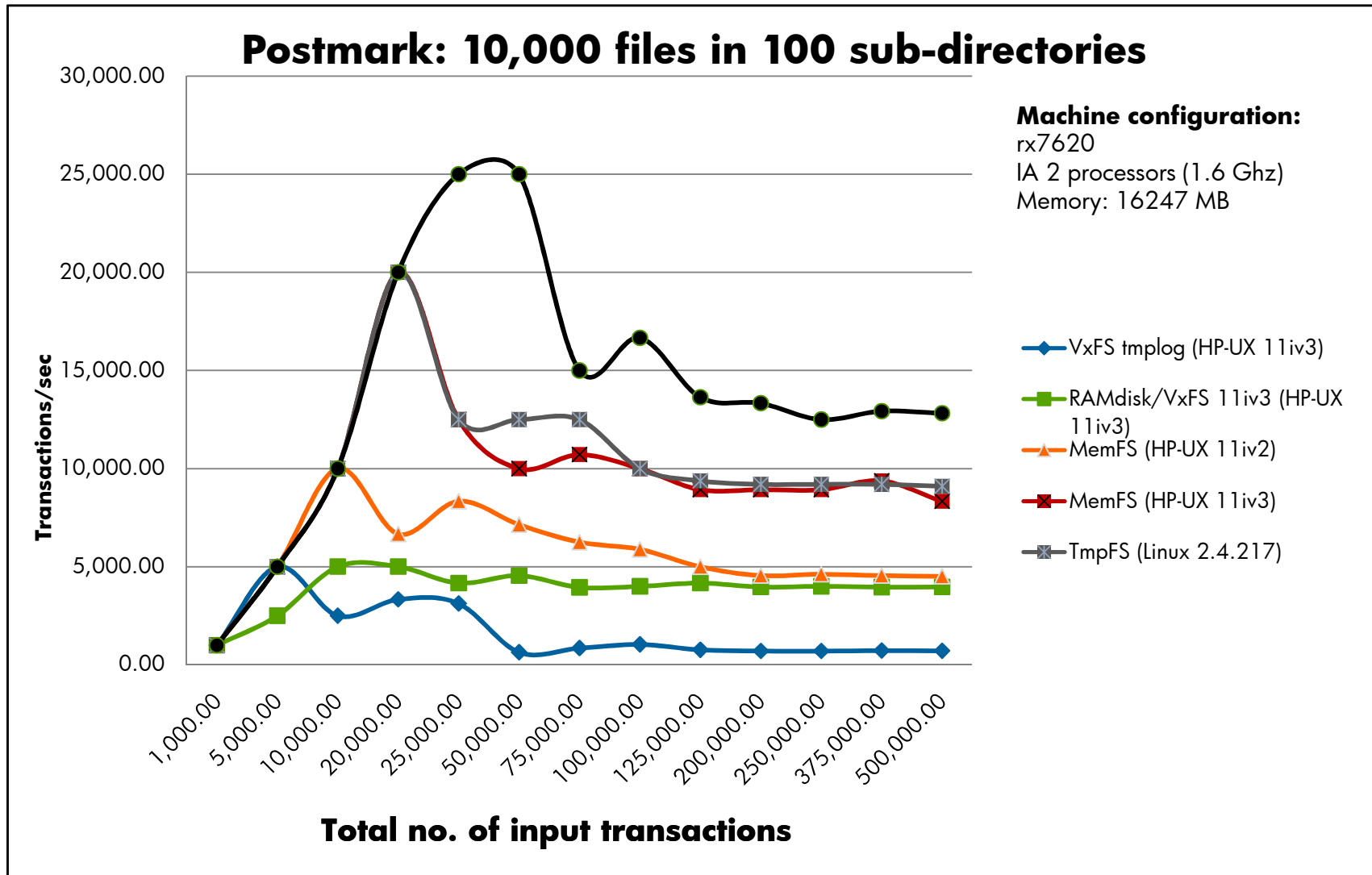
# Competitive data (continued)

Feature	HP-UX 11i v3 MemFS	HP-UX 11i v2 MemFS	Tmpfs (Solaris & Linux)	MFS (Tru64 UNIX)	HP-UX RAMdisk
Metadata Handling - footprint	Limited to percentage of physical memory. Non Swappable	Limited to a percentage of BC/FC. Non-swappable	Limited to percentage of physical memory. Non Swappable.	Limited to percentage of physical memory. Swappable.	Limited to percentage of physical memory. Swappable.
Data Handling - Swapping	Under UFC and system memory pressure.	Under BC and system memory pressure. Additional pressure by process swapping.	Under system memory pressure only	Under system memory pressure. Process. Additional pressure by process swapping	Under system memory pressure only. Additional pressure by process swapping
High performance limit	Linked to physical memory size	Linked to physical memory size	Linked to physical memory size	Linked to physical memory size	Linked to physical memory size
Min Virtual Memory usage (VM=>PM +Swap)	Metadata size	Metadata size	Metadata size	Total file system size	Total file system size

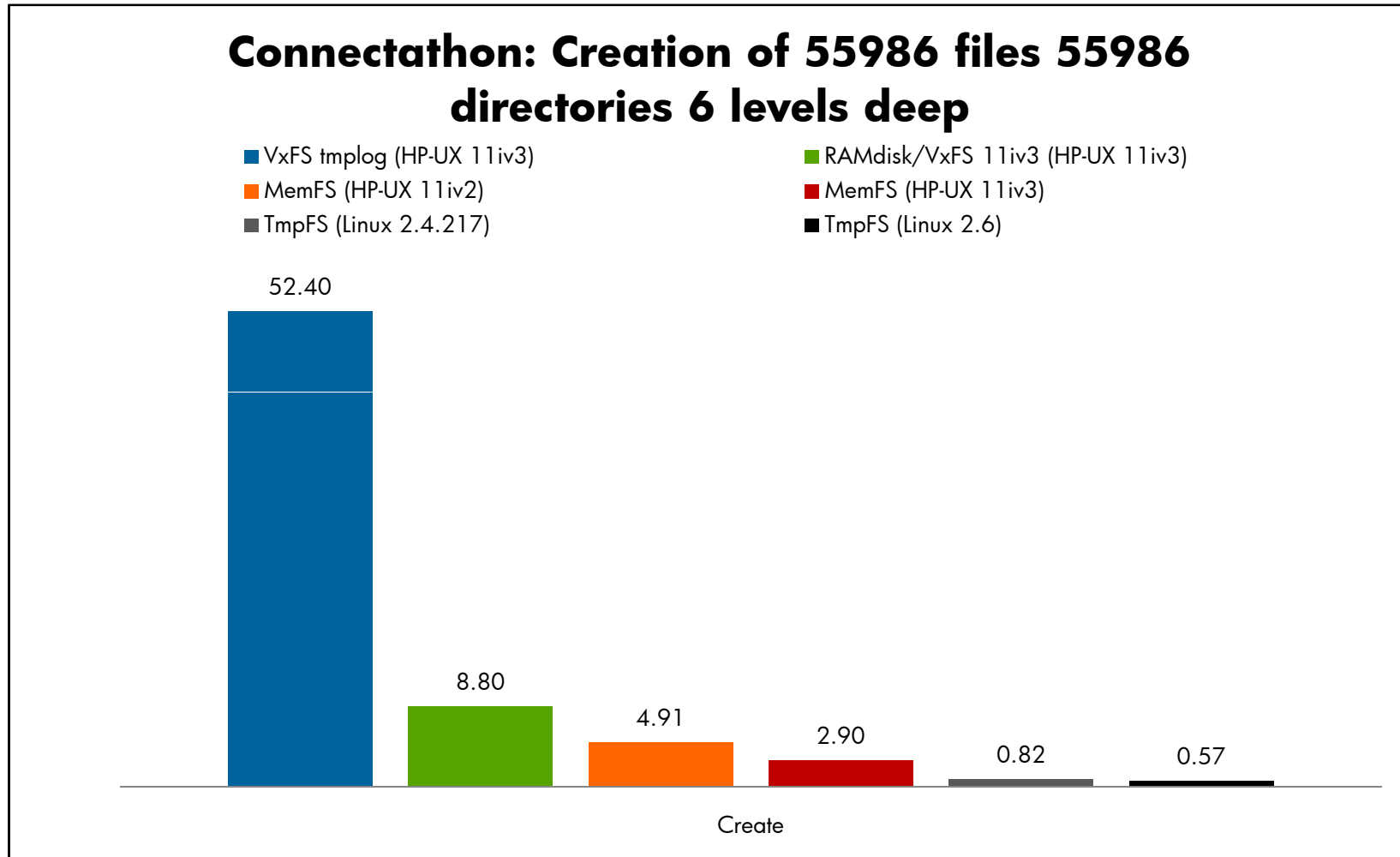
# Competitive data (continued)

Feature	HP-UX 11i v3 MemFS	HP-UX 11i v2 MemFS	Tmpfs (Solaris & Linux)	MFS (Tru64 UNIX)	HP-UX RAMdisk
Worst Case impact on system memory	Occupies major portion of UFC. All file system operations slowed down.	Occupies major portion of BC/FC. All file system operations slowed down.	Occupies major portion of virtual memory. System can come to halt.	Occupies major portion of virtual memory. System can come to a halt.	Occupies major portion of virtual memory. System can come to a halt.
Maximum number of instances	Limited by available Physical memory	64(can be tuned)	Not available. Probably unlimited	Not available. Probably unlimited	15
Maximum File Size	Limited by virtual memory and swap size	256 GB	Limited by virtual memory and swap size	1 TB (MFS uses UFS as base)	Depends on FS and limited by data segment size
Maximum Filesystem Size	Limited by virtual memory and swap size	256 GB	Limited by virtual memory and swap size	1 TB (MFS uses UFS as base)	Limited by data segment size

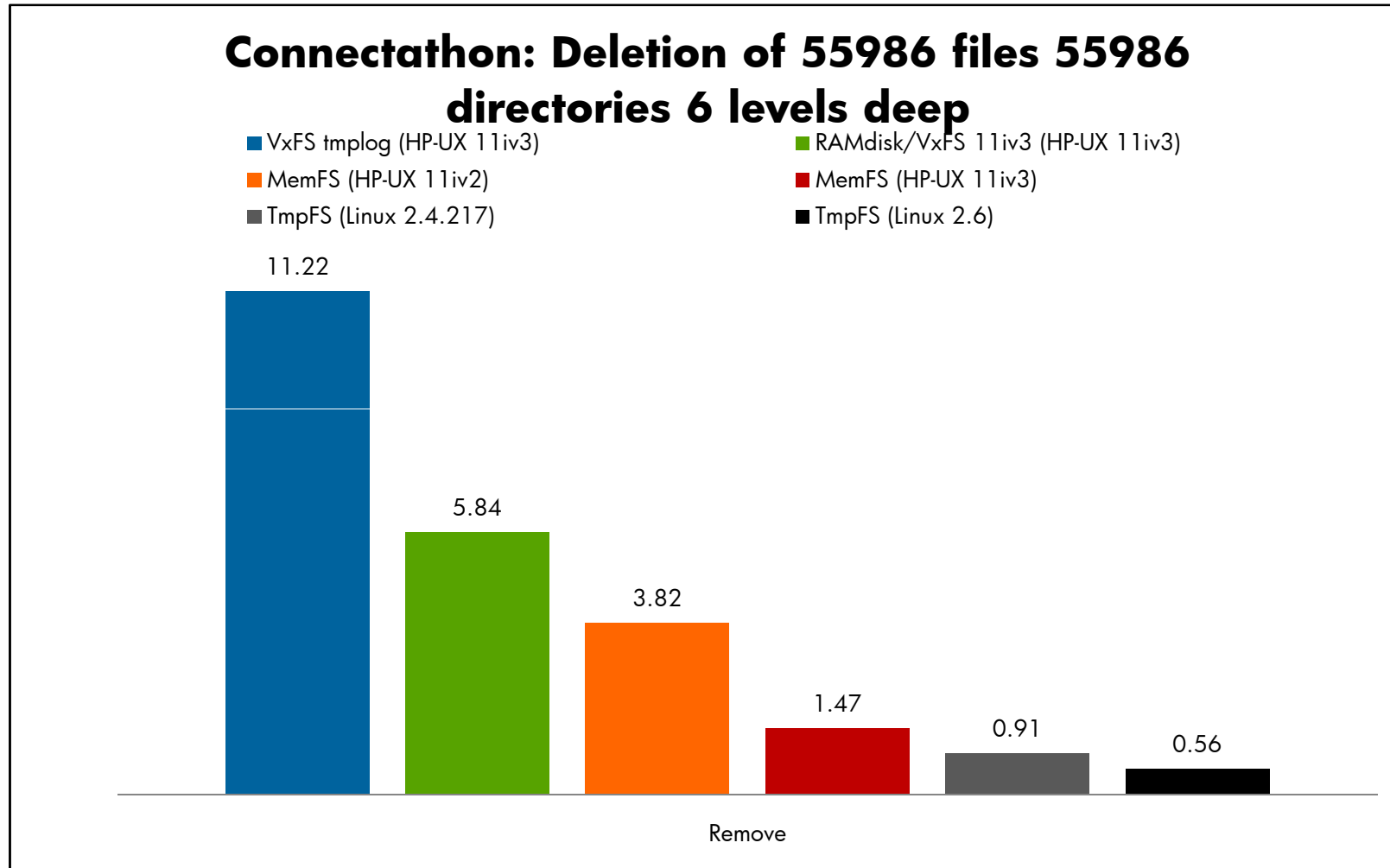
# Performance comparisons (Postmark)



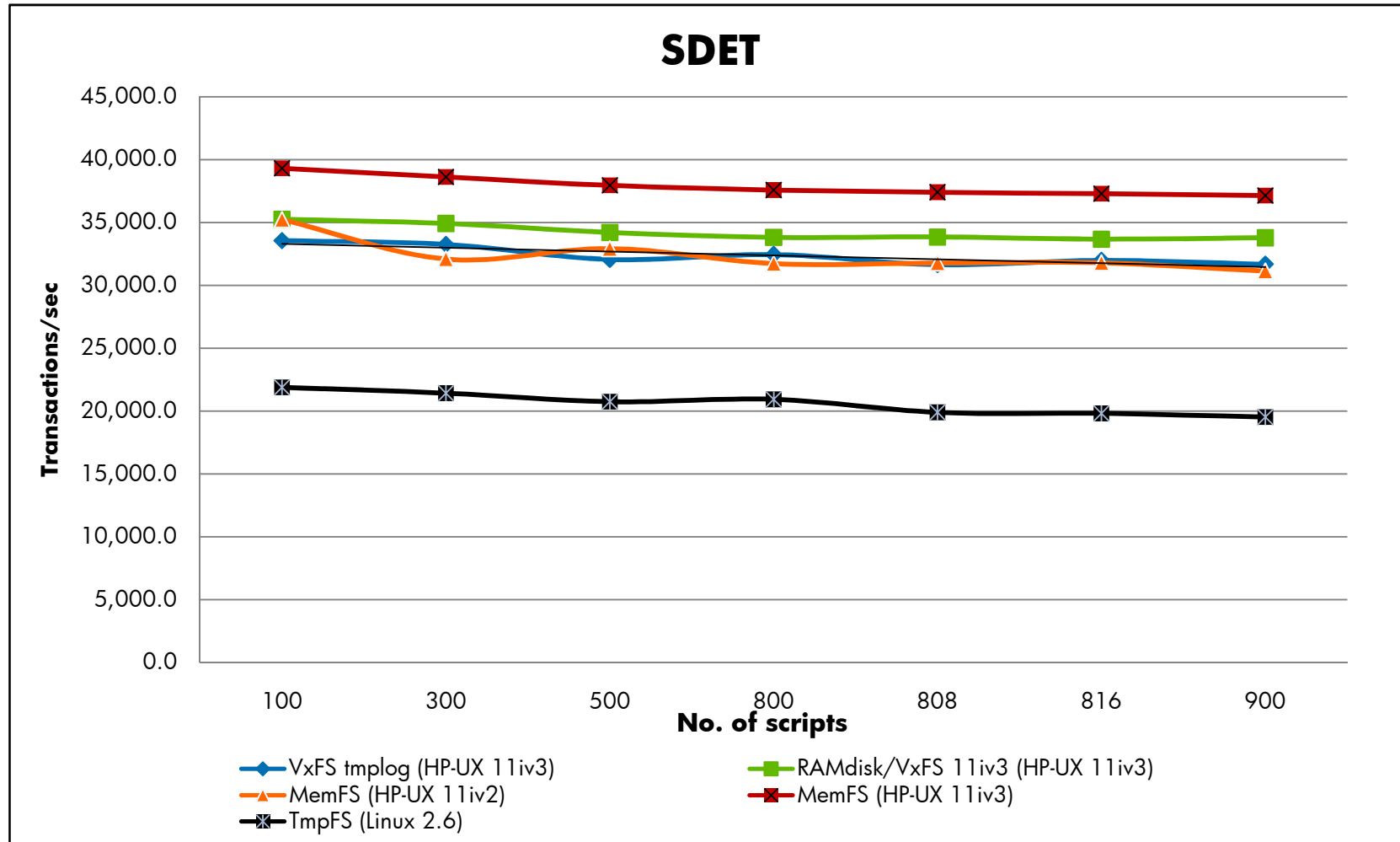
# Performance comparisons (Connectathon)



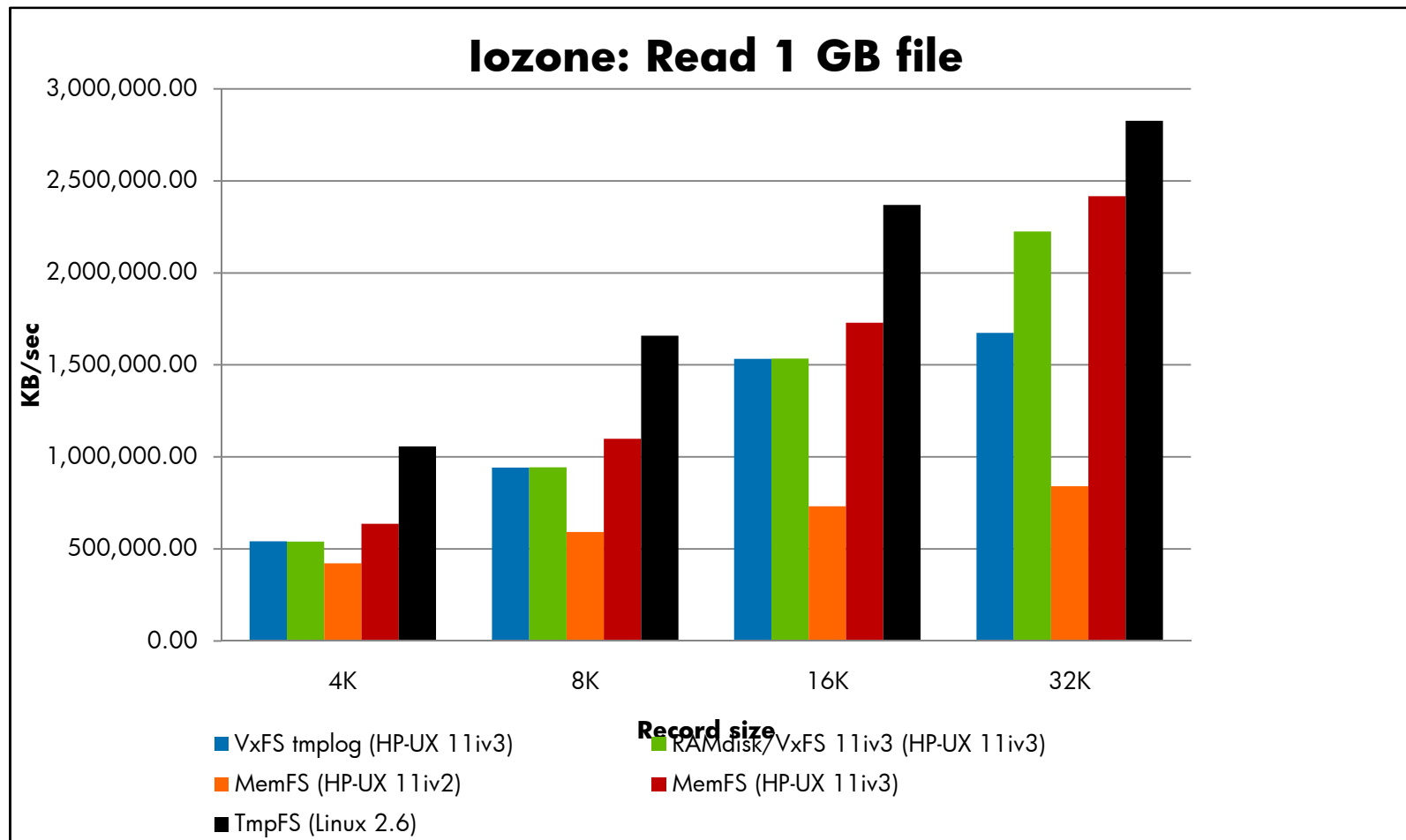
# Performance comparisons (Connectathon)



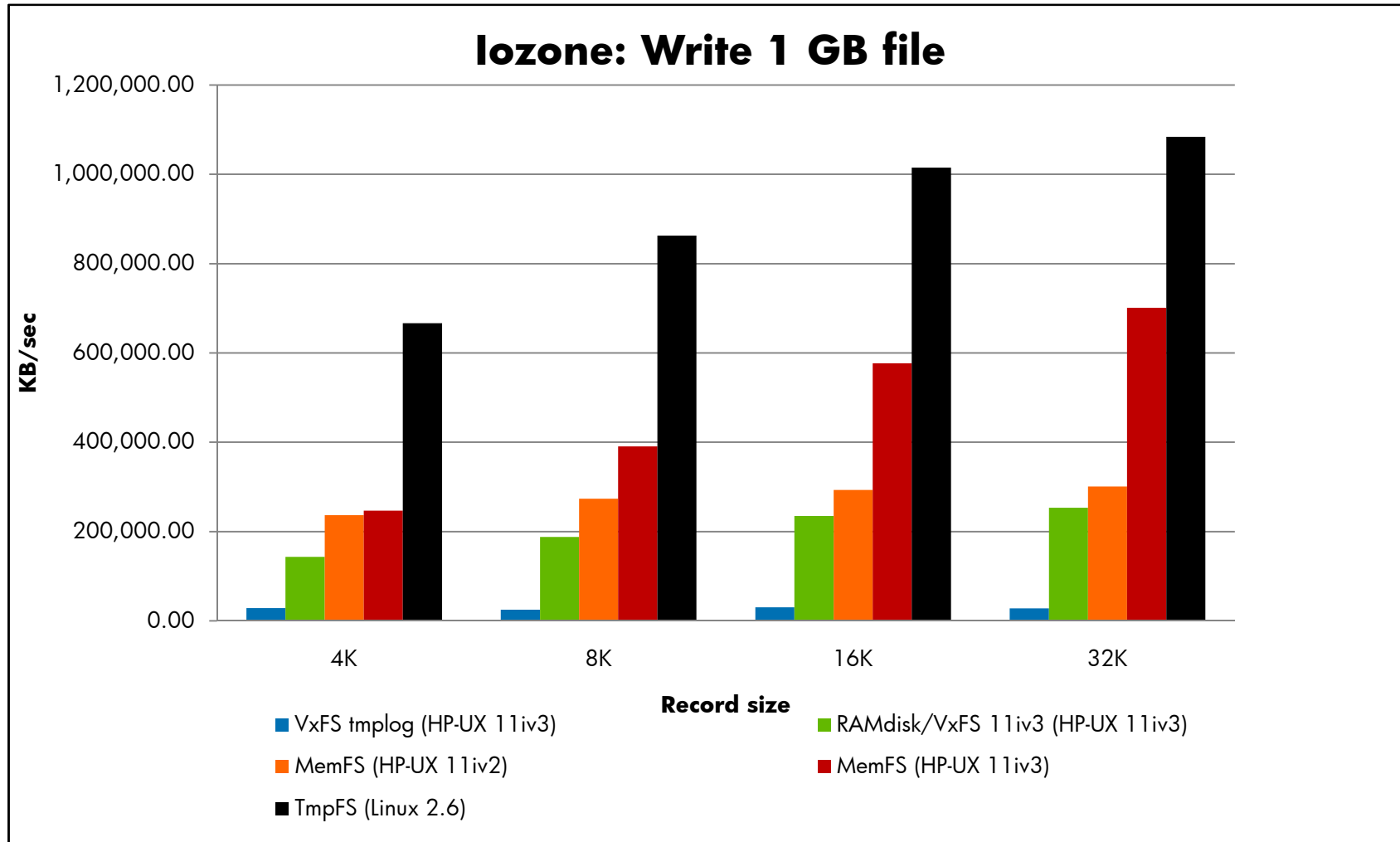
# Performance comparisons (SDET)



# Performance comparisons (iozone)



# Performance comparisons (lozone)



# Installing, configuring and tuning MemFS



# Installing and configuring MemFS

- Delivery model: EP/NCF
  - Web release
  - AR0903 release
- MemFS includes
  - 6 EPs, 1 NCF
- EP patch numbers:
  - PHCO\_38751, PHCO\_38752, PHCO\_39038, PHCO\_39039, PHKL\_38077, PHKL\_38949
- Installing MemFS:

```
swinstall -x autoreboot=true -s <depot-path> MemFS
```
- Removing MemFS:

```
swremove -x autoreboot=true MemFS
```



# Product deliverables

- Files delivered

- /usr/conf/mod/memfs MemFS static module
- /sbin/fs/memfs/mount MemFS mount command
- /sbin/fs/memfs/umount MemFS umount command
- /var/adm/sw/update\_prep/MemoryFSKern1131.100  
Update-UX script

- To check whether MemFS is loaded or not:

- `kcmodule -v memfs`

State of the module must be static

# MemFS on HP-UX 11i v3: Summary of features

- A new light weight layout: Performs better than MemFS on HP-UX 11i v2 by **a factor of 2**
- Support for large file and file system sizes (Currently tested up to 500GB)
- No limitation on the number of MemFS instances
- Number of files on a file system can be restricted, if needed
- File system size can be restricted or can be allowed to grow depending on the swap space available
- Swaps data on to system swap device efficiently
- Support for large files
- Maximum amount of memory and swap that can be used by MemFS can be tuned
- Access Control Lists (ACLs) are not supported

# Creating a MemFS file system

- mkfs is not needed
  - `mount -F memfs [[-o remount] [-o size=<size>KB/MB/GB][ -o ninode=<xxx>]] /<mount_point`
    - size: Maximum size of the file system. Won't assure the size.
    - ninode: Maximum number of inodes on the file system.
    - remount: To modify these options
- Example:
  - To mount a file system of maximum size 100MB
    - `# mount -F memfs -o size=100MB /tmp`
  - To remove the restriction on maximum number of inodes of /tmp
    - `# mount -F memfs -o remount,ninode=0 /tmp`
- Note: if new values specified is less than current usage, following warnings are displayed by command:
  - `mount: Warning! current number of files exceeds the ninode specified`

# /etc/fstab entries

- To create MemFS instance automatically during boot, add entry in /etc/fstab
  - memfs            directory    memfs            size=<size> 0 0            #comment

- The first field is always ignored – no matter even if it be a valid device name

- An fstab entry of type memfs will be considered only if the mount point (directory) matches the directory argument specified for mount command

- For eg:

- In /etc/fstab

```
/dev/vg00/lvol8 /memfs memfs defaults 0 0
/dev/vg00/lvol8 /var vxfs delaylog 0 2
```

```
# mount /dev/vg00/lvol8
```

```
mount: /dev/vg00/lvol8 is already mounted on /var
```

```
# mount /memfs
```

```
#
```

# /etc/mnttab entries

- Mounted MemFS filesystems will appear in /etc/mnttab with the device field "memfs"
  - Example:
    - # cat /etc/mnttab
    - memfs /memfs memfs largefiles,size=2gb,dev=1000004 0 0 1227266552
    - # mount -v
    - memfs on /memfs type memfs largefiles,size=2gb,dev=1000004 on Fri Nov 21 16:52:32 2008
- The device name is not unique for a memfs instance: instance is always "memfs"
- The MemFS instance can be unmounted only by specifying the mount point and not using "memfs" (even if there is only one mounted MemFS instance)
  - # umount memfs
  - umount: cannot find memfs in /etc/mnttab
  - cannot unmount memfs
  - # umount /memfs

# MemFS tunables

- `memfs_metamax`

- Specifies the maximum amount of kernel memory that can be used by MemFS. It doesn't reserve/assure the memory for MemFS.
- Ranges from 1-30% of kernel memory. Default is 15%
- Can be specified as: Default, percentage value (10%) or constant (400MB)
  - `kctune -s memfs_metamax=400M`
- If MemFS utilizes all the memory specified by this tunable, mounting of new MemFS file systems and creation of new files on them will fail and following message will be displayed on console:

```
memfs: metadata exceeds memfs_metamax
```

- If specified value is less than the memory currently is use, following warning is displayed on console

```
Warning: The specified value of tunable memfs_metamax is less than the amount of memory currently in use by MemFS, 20480 bytes
```

- OL\* aware

# MemFS tunables

- `memfs_swapmax_pct`
  - Specifies the maximum amount of swap that can be used by MemFS. It doesn't reserve/assure the space for MemFS.
  - Ranges from 0-80% of kernel memory. Default is 50%
  - Can be specified as: Default or percentage value (10%)
    - `kctune -s memfs_metamax=10%`
  - If MemFS utilizes all the swap space specified by this tunable, adding new data or extending the contents of the files will fail:
    - `memfs: Cumulative data exceeds memfs_swapmax_pct`
  - If specified value is less than the swap space currently is use, following warning is displayed on console
    - Warning: The specified value of tunable `memfs_swapmax_pct` is less than the percentage of swap space currently in use by MemFS, 15 percent
  - OL\* aware

# Tools support

- **sar**
- A new product on sar is released to provide MemFS details (SAR\_MEMFS\_ENH)
- -z option reports MemFS details
- MemFS details can not be recorded

```
# sar -z 1 10
```

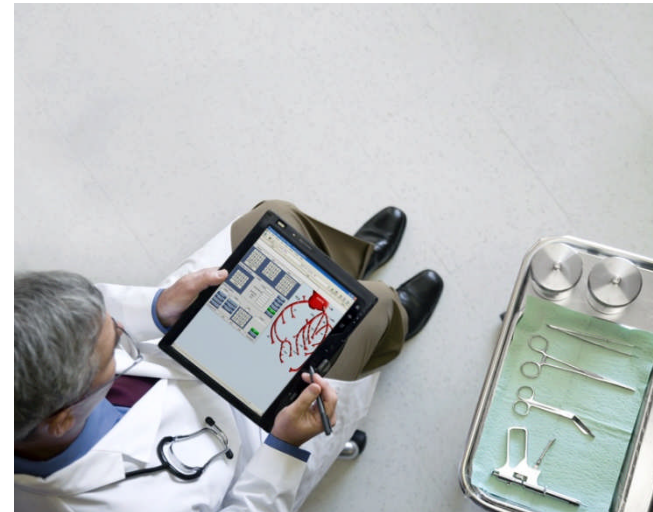
```
HP-UX librx262 B.11.31 U ia64 11/26/08
```

```
15:39:59 blkcnt swpcnt
15:40:00 30031 0
15:40:01 30031 0
15:40:02 30031 0
```

blkcnt: Number of dirty pages in memory/UFC

swpcnt: Number of dirty pages in Swap

Above counters are in terms of 4K only and  
not in base\_pagesize



# GlancePlus memory report – MemFS details

```

Glance C.04.70.000          15:43:30 librx262      ia64      Current  Avg  High
-----
CPU Util  SSA
Disk Util FF
Mem Util  S SU U
Swap Util U UR R
-----
MEMORY REPORT
Users= 2
Event      Current  Cumulative  Current Rate  Cum Rate  High Rate
-----
Page Faults      0          178          0.0          40.4        230.0
Page In           0           0           0.0           0.0         24.2
Page Out          0           0           0.0           0.0          0.0
KB Paged In      0kb         0kb          0.0           0.0          0.0
KB Paged Out     0kb         0kb          0.0           0.0          0.0
Reactivations    0           0           0.0           0.0          0.0
Deactivations    0           0           0.0           0.0          0.0
KB Deactivated   0kb         0kb          0.0           0.0          0.0
VM Reads         0           0           0.0           0.0          0.0
VM Writes        0           0           0.0           0.0          0.0
Total VM :      605mb  Sys Mem : 856mb  User Mem: 142mb  Phys Mem : 2.0gb
Active VM:      476mb  Buf Cache: 0mb  Free Mem: 816mb  FileCache: 222mb
MemFS Blk Cnt: 30031  MemFS Swp Cnt: 0  Page 1 of 1
ProcList CPU Rpt Mem Rpt Disk Rpt NextKeys SlctProc Help Exit
  
```

GlancePlus C.04.70.000 reports MemFS details (Memory Report)

# Unsupported features

- Access Control Lists (ACL) are not supported
- Quotas are not supported
- Creation of empty directories
- Hard link to directories

# Guidelines for using MemFS

- Use only for temporary files
- Do not use /tmp & /var/tmp for MemFS
- Do not allow MemFS to occupy full memory



# Documentation

- White paper

<http://docs.hp.com/en/5992-5789/5992-5789.pdf>

- Administrator's guide

<http://docs.hp.com/en/5992-5788/5992-5788.pdf>



Technology for better business outcomes

