Unix Performance Fundamentals

By: Jeff Kubler

Kubler Consulting

jrkubler@proaxis.com

www.proaxis.com/~jrkubler

Phone: 541-745-7457

What is Performance?

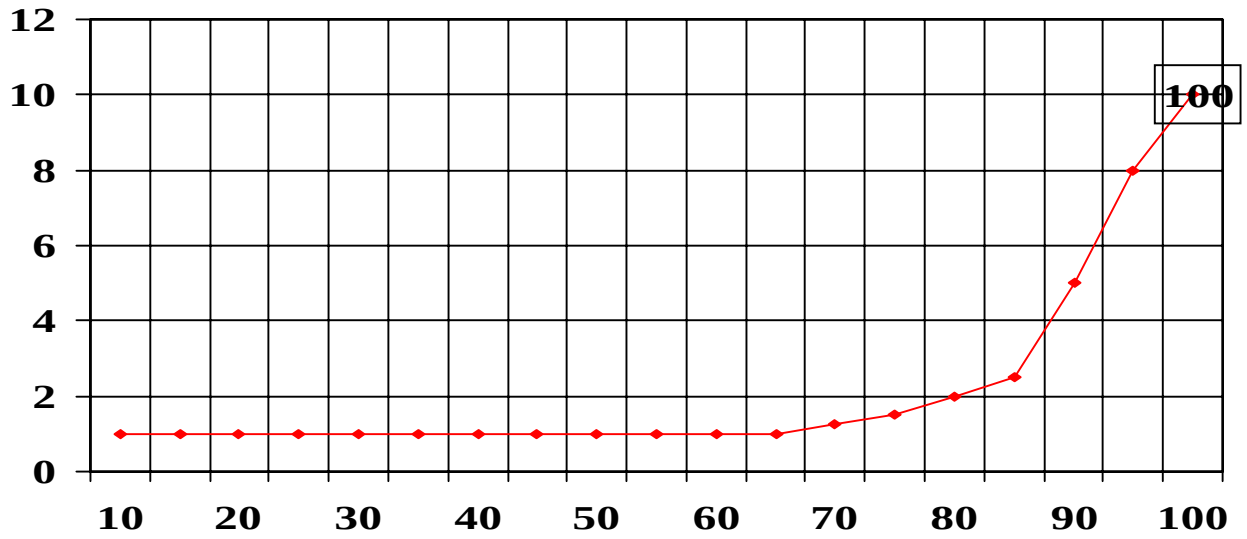
- The execution of an action, something accomplished, the fulfillment of a request, ….
- Good performance -- when requested actions complete within expectations. Bad is when they don’t!
- Performance is important because users have expectations!

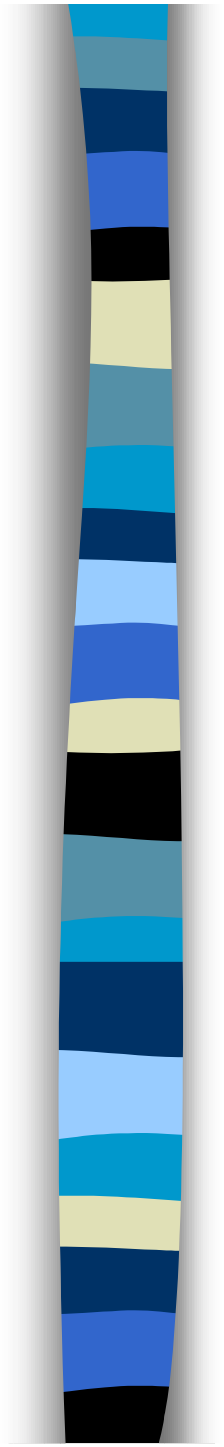
Introduction

- Four basic macro areas: CPU, Memory, disk, network. Several micro areas: database, application.
- Performance Management: keeping resources adequate
- Resource restrictions are often called “bottlenecks”
- Bottlenecks are first encountered at the “Knee in the Resource Curve”

Basics

- Importance of response times.
- Avoid resource overuse or bottleneck
- Bottlenecks often called “Knee in Curve”





Performance -- General Terms

- Resources often measured as a percentage of a whole
- Sometimes a count is the important measurement (example: the count of processes waiting in a queue)
- Averages are often the important indicator
- Performance may be acceptable until one more “straw” is added.

Macro Area: CPU

- CPU Activity
- CPU Activity States
  - Busy
    - Useful Work
  - Idle
    - CPU in the “Bank”
  - Paused for I/O
    - Wait for Disk
- CPU Run/Ready Queue

Macro Area: CPU

- Process activity handled by the scheduler
- Scheduler has set time slice of 1/10 of a second
- Processes assigned either Real Time or Time Share
- Priorities calculate dynamically

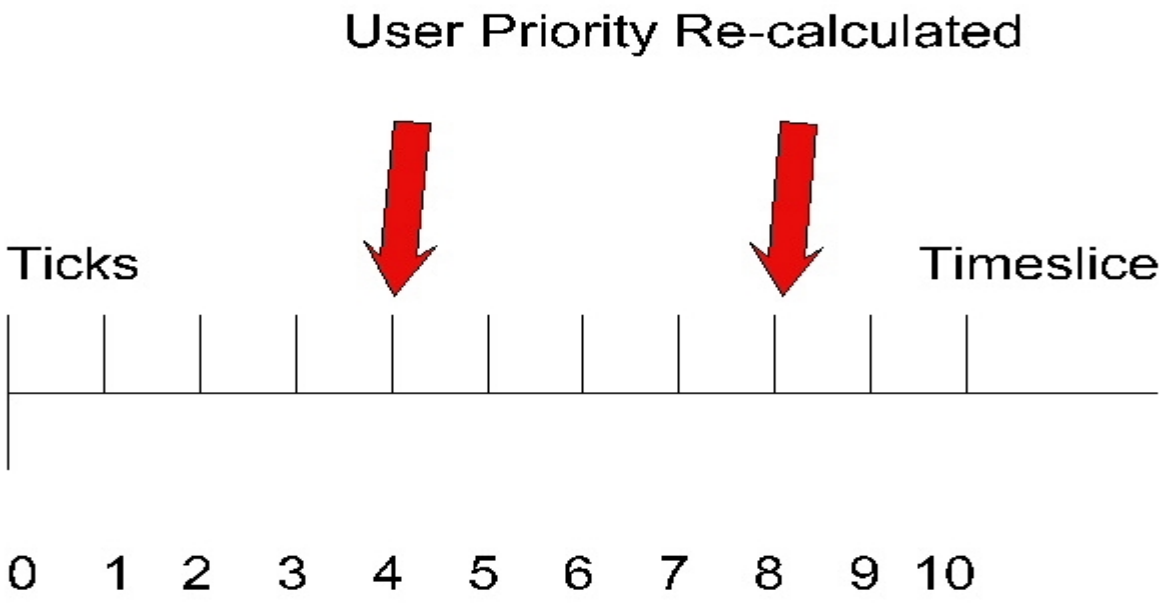
CPU: Measurement Metrics

- CPU:
  - Total Busy: the resource as 0 to 100 percent utilized.
  - Run Queue: the number of processes awaiting the CPU.
  - Total Busy makeup: User, Real, Nice, Nnice, System, Interrupt, Context Switching, Trap, Mem, and Idle. Capture Ratio helps understand CPU usage.

CPU: Measurement Metrics - Scheduler's Rules

- (R)Real Time xecute at highest priority.
  - Import system proc
  - preempt low priority
  - runs until sleeps
- Time Share are time sliced 1/10 sec.
  - Made up of sys&user
  - higher pri proc can preempt
  - Round Robin scheduling
  - Priorities are recalculated dynamically
  - Nice value can be used to adjust priority calculations
  - statdeamon recalculates process priorities

CPU: Time Slice



CPU: HP-UX Process Priority Calculation

newpri=(“recent CPU”/const) + basepri + nice)

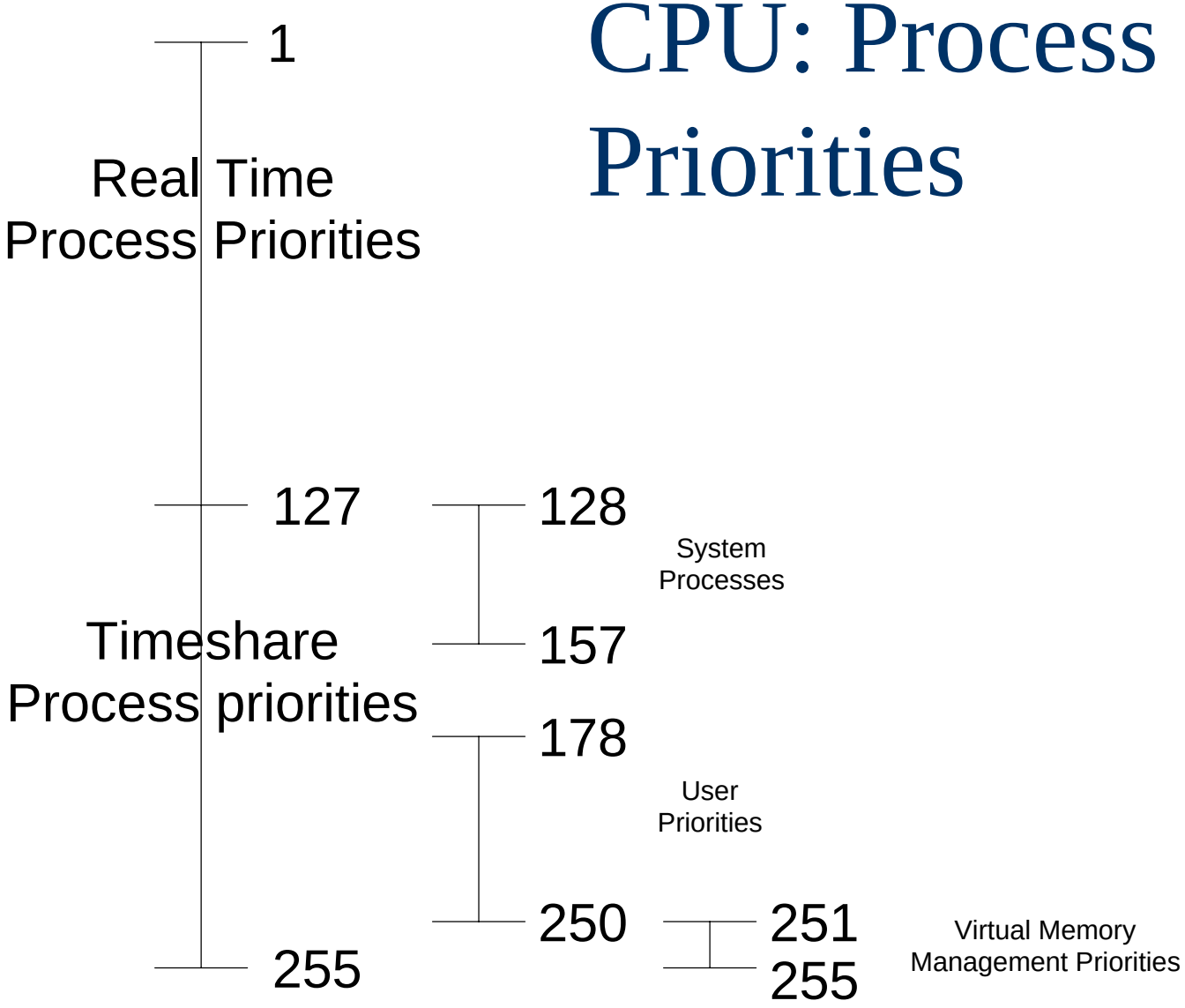
```
PID   Name           User Name TTY      CPU% Nic Pri  RSS/Size …..
**********************  Workload:DETACH Processes  **********...
16577 smx0020        ptman      ---      27.0  20 154 3628/  9m ...
16590 oracleptmn     ptman      ---      23.5  20 154 131m/137m1...
14895 oracleptmn     oracle     ---       6.9  20 154 131m/137m ...
16647 oracleptmn     ptman      ---       5.1  20 148 131m/137m7...
15720 oracleptmn     oracle     ---       4.5  20 154 131m/137m7...
16315 ldupt          root       ---       3.8   R  48 3536/4044 ...
```

CPU use

Newpri

```
TTY   PID USERNAME  PRI NI
SIZE   RES  STATE   TIME %WCPU  %CPU COMMAND
   ?  1730 root     235 20   448K   184K run   420:18 65.89  ...
```

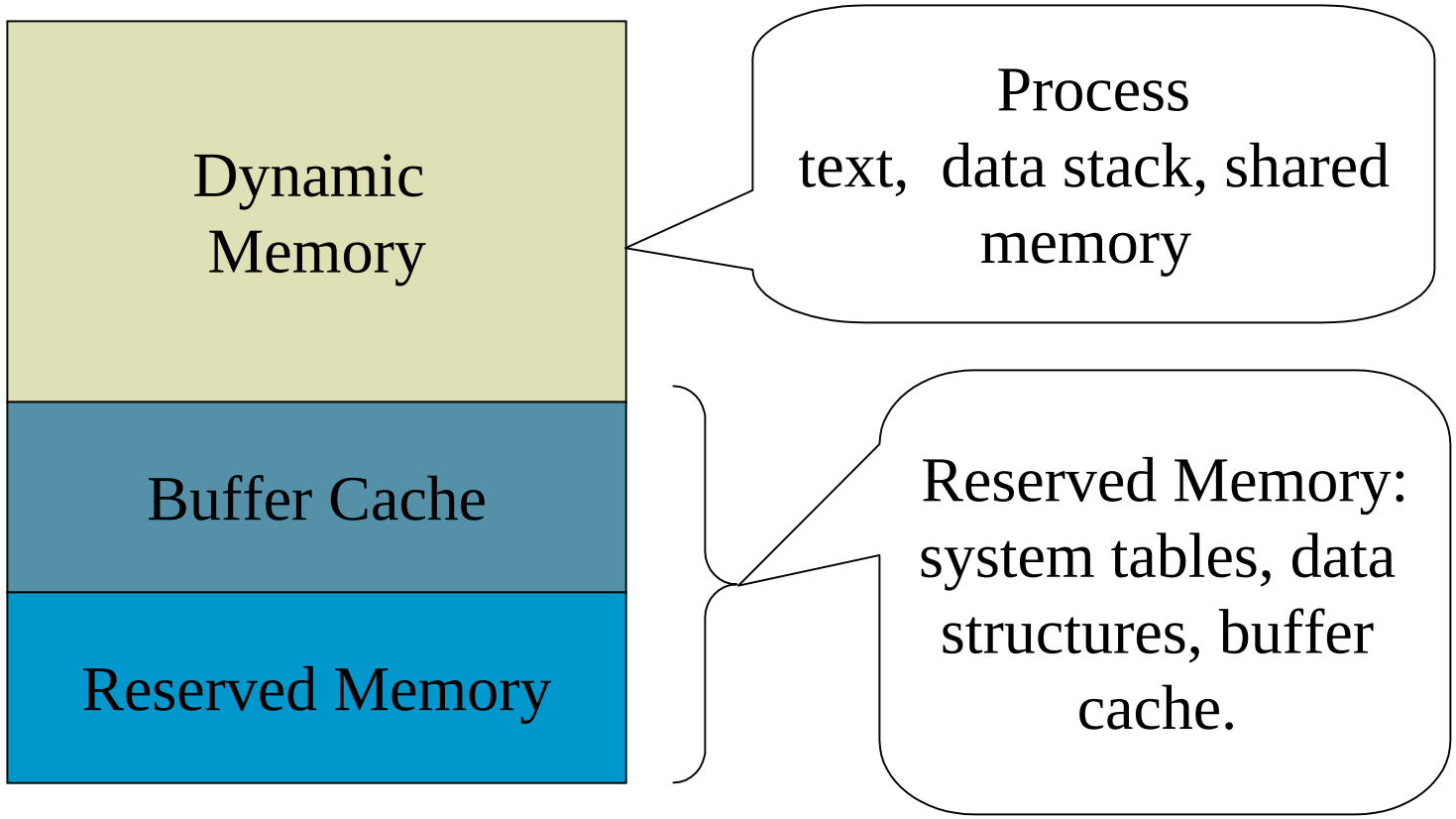
CPU: Process Priorities



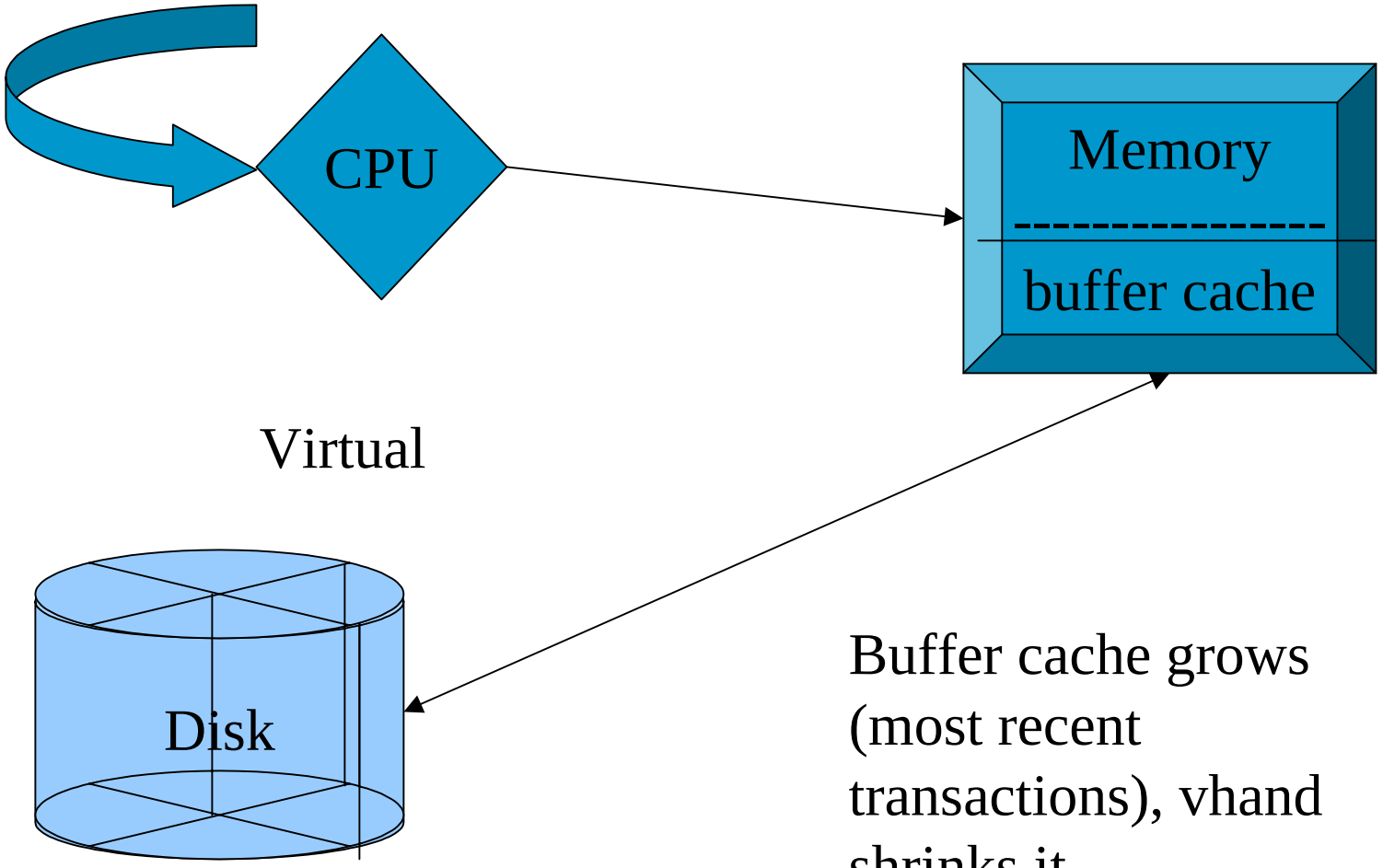
Macro Area: Memory

- Virtual Memory/Swap
  - Inexpensive means to make Main Memory look big!
- Paging - handled by the vhand process
- Swap (<=9.04) versus = Deactivation (>=10.0)
- 11.0 Supports Variable-Sized Pages (Better Performance)

Memory: Allocation and Use



Memory: buffering and virtual

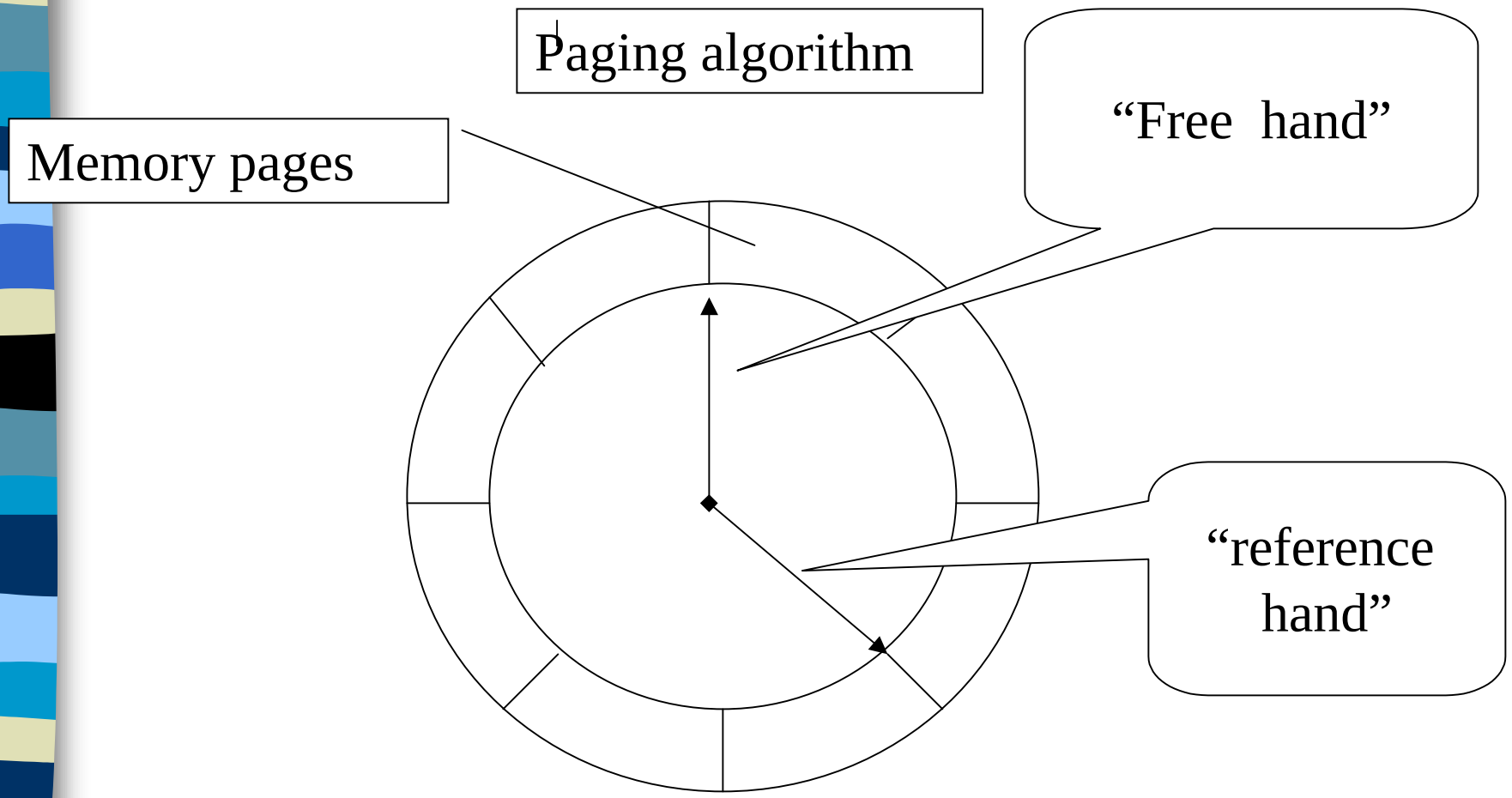


Buffer cache grows (most recent transactions), vhand shrinks it,

Memory: VHAND Processes

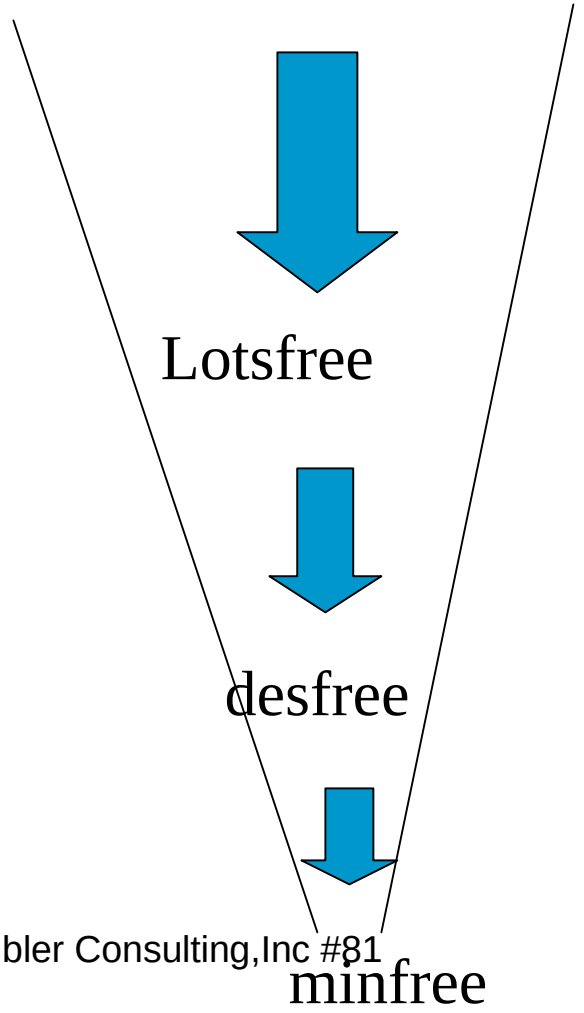
- paging - a technique which takes the pages (virtual memory address) from memory and moves them to virtual memory (on disk). This is called by the VHAND process.
- page fault - and invalid address. This occurs when the CPU requests a page from memory and it does not find the page.
- Deactivation - moving a process from the swap space on disk to main memory and from main memory to disk. This is done by the swapdaemon.
- thrashing - when a process spends more time paging than processing. A high number of page faults.

Memory - vhand process



Memory: Paging and Deactivation Thresholds

- ***Lotsfree*** **- Upper threshold at which paging daemon *vhand* will begin stealing pages**
- ***desfree*** **- "Desired Free" - Panic button for Swapper- lower threshold for *vhand***
- ***minfree*** **- Least amount of memory that is tolerable before the system begins process deactivation (sys proc *sched*)**



Memory: Measurement Metrics

- Memory Analysis (scratch pad for work):
  - Memory Used %, Virtual Memory %.
  - Paging (in and out), Deactivations
  - Read and write buffer cache hits.

Macro Area: Disk

- Disk I/O (long term storage):
- Virtual Memory:allows programs with memory requirements > than memory to load.
- Swap Space:function that moves pages to disk and back.

Disk: Measurement Metrics

- Metrics:
  - Disk I/O Queue Length - requests waiting for service.
  - Total I/O - total reads and writes.
  - Disk service time - % of time a device is used.
  - Read/Write hit buffer cache efficiency - how many I/O's eliminated due to buffer

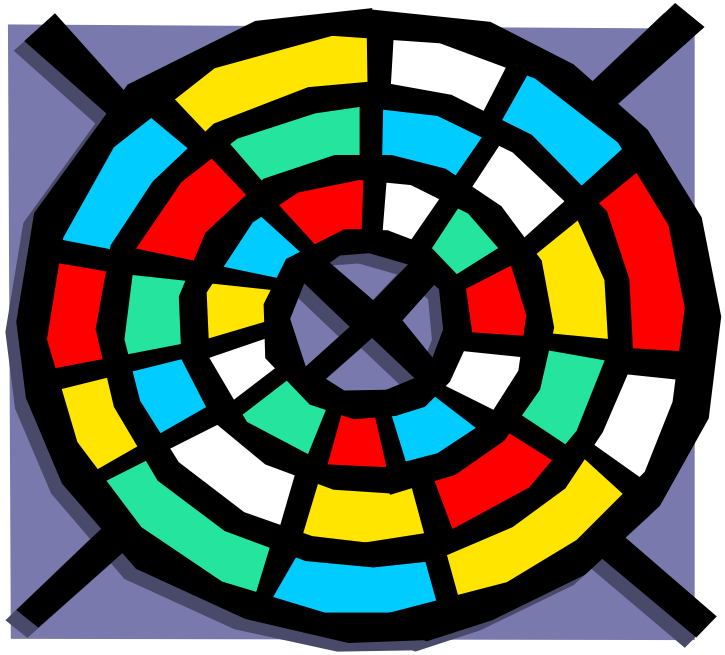
Disk: Data Locality

- Describes the location of data on disk (it is sometimes referred to as locality of reference)
- Data Locality encompasses both the issue of the placement of files on disk or on multiple disks and the issue of records within the files placed on disk.

Disk: What is Disk I/O?

- Act of retrieving and/or updating information stored on a disk drive or in a disk environment.

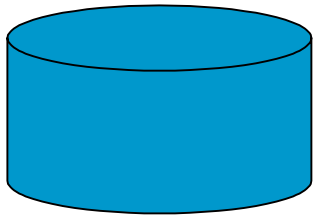
Overhead - Negotiating the controller.
Seek Time - find data
Latency - wait for data spin.
Xfr (transfer of data) - bring data over.



Disk: Anatomy of a Process

- All activity exists a process.
- Processes usually rely on data. Data in one of two places, in memory or on disk.
- If on disk then if updated it must be posted back to disk.
- Disk access is the slowest link.

Disk: Problem I/O Issues



- Disk I/O Imbalance
- Hardware
- Configuration issues
- Disk and File Fragmentation
- File system choices and configuration (default values in JFS favor data integrity).
- Database inefficiencies
- Buffer cache configuration and use
  - not enough causes additional I/O
  - too much and users memory storage suffers
- Stripping can become fragmentation.

General Measurements of Disk I/O

- Disk I/O Queue Length
- Pause or Wait for I/O
- Disk Service time
- Disk Utilization
- Total I/O count
- Buffer Cache efficiency
- Response times

Macro Area: Network

- Network (data transfer, NFS activity):
- Network traffic in packets.
- Problems:
  - Poorly planned and overtaxed networks
  - shared files on network file systems
  - inadequate hardware

Network: Measurement Metrics

- Packets in/Packets out
- Errors in/Errors out
- % of Collisions
- Count of errors

Micro: Application Analysis

- Use workload groupings when possible. Glance, SOS, etc. provide these.
- Look at individual processes. What files do they open? What wait states do they encounter?

Relational Database inefficiencies

- **Example: ORACLE, INGRESS, INFORMIX, PROGRESS**
- **Consist of: Tables, Indexes, Rollback logs, and Before Image Logs**
  - **Suggestions:**
    - Optimize placement of Tables and Indexes.Place table files, indexes, and logs on separate disk drives.
    - Use supplied optimization tools

Strategies

- **Memory**
- **Buffer Cache**
- **JBOD**
  - **balance I/O, work on fragmentation,**
- **Stripping**
- **Raw I/O Vs. File System I/O**

HP-UX Memory - How Much is Enough?

Kernel minimum - 7 MB

Basic OS Utilities - 10 MB

Minimum total for a few light users - 16 MB

MB per heavy users - 2 to 5 MB

X users (per user) - 10 MB

RDBMS app (server) - 10 to 25 MB

Each RDBMS user (clients) - 1 to 3 MB

Configuration issues in the Disk Environment

- make sure you have enough -- use bdf
- avoid too many devices per channel
  - How many is too many? General rule: avoid more than 10 fast & wide per card
  - Look to any third party providers for additional info
- Too small or inappropriately placed swap space or buffer size.

Configuration Issues -CPU

- CPU's delivered in "Families", relative performance rates them in respect to members of the family and others.
- Multi-processors allow easy scalability
- The MP effect means that additional CPU's do not give the same impact as you increase

Unix Onboard commands

- CPU
  - Top - displays top 10 CPU users
  - sar - gathers info on system use
  - ps (-ef) - displays process info
  - uptime - gives load averages
- Memory
  - vmstat (-n -s) - reports virtual memory info
  - iostat (-t) - displays I/O rate for disks

Unix Onboard commands

- Disk
  - ioscan - displays I/O devices
  - bdf - report free disk blocks
  - swapinfo - system paging info
  - LVM commands: vgscan, vgdisplay, etc.
- Network
  - netstat (-s) - display network status
  - lanscan - lan device config and status
  - lanadmin - lan administrator

Top Command

```
Load averages: 2.31, 1.74, 1.54
102 processes: 100 sleeping, 2 running
Cpu states:
 LOAD   USER   NICE    SYS   IDLE  BLOCK  SWAIT   INTR   SSYS
 2.31  90.8%   0.0%   9.2%   0.0%   0.0%   0.0%   0.0%   0.0%

Memory: 8272K (5988K) real, 19100K (14528K) virtual, 2184K free  Page# 1/8

 TTY   PID USERNAME  PRI NI   SIZE   RES  STATE   TIME %WCPU  %CPU COMMAND
   ?  1730 root      235 20   448K   184K run   420:18 65.89 65.78 _mprosrv
   ?  4615 root      168 20  1272K   520K sleep 245:01 25.25 25.20 _progres
   ?  1563 root      128 20   480K   216K sleep 432:17  2.04  2.04 _mprosrv
   ?  6187 root       48  0  1732K  1788K sleep   2:30  1.01  1.01 lpsmid
   ?   961 root      154 20     8K    16K sleep 188:53  0.93  0.93 nfsd
This process services NFS requests from remote systems.
   ?   962 root      154 20     8K    16K sleep 185:05  0.91  0.91 nfsd
   ?   956 root      154 20    32K    40K sleep 185:28  0.91  0.90 nfsd
   ?   957 root      154 20     8K    16K sleep 187:26  0.90  0.90 nfsd
   ?     7 root      -32 20     0K     0K sleep  67:26  0.32  0.32 ttisr
d1p0  6309 root      178 20   208K   308K run     0:00  0.33  0.29 _mprosrv
   ?     0 root      127 20     0K     0K sleep  71:35  0.21  0.21 swapper
   ?  1407 root      154 20   180K     0K sleep  71:35  0.22  0.22 swapper
   ?     2 root      128 20     0K     0K sleep  18:55  0.21  0.21 vhand
   ?  1411 root      156 20   180K    68K sleep   6:32  0.15  0.15 _mprshut
```

Sar (- u CPU, -b buffer, -d device, etc.)

```
# sar -u 1 10

HP-UX pointman B.10.20 E 9000/831    03/01/99

21:09:44    %usr    %sys    %wio   %idle
21:09:45       1       0       0      99
21:09:46       0       0       0     100
21:09:47       0       0       0     100
21:09:48      18       0       0      82
21:09:49       0       3       0      97
21:09:50       0       0       0     100
21:09:51       0       0       1      99
21:09:52       0       0       0     100
21:09:53       0       0       0     100
21:09:54       0       4       0      96
Average        2       1       0      97
```

Measurement of Disk I/O - vmstat

```
Vmstat -d 5
|procs|  |-memory-| |--------page------|  |-faults-|  |-cpu--|
r  b  w  avm   free re at pi po fr de sr  in  sy  cs  us sy id
1 46  0 2469    466  0  0  0  0  0  0  0 108  37  25   3  2 95
0 47  0 2140    500  1  1  0  0  0  0  0 113  65  30   2  1 97
device    xfer/sec
c0t6d0           0
c0t1d0           0
```

- Procs: Running, Blocked, Swapped
- Memory: Active Virtual Pages; size of memory free
- re: Re-claims; Page Freed but Referenced Again
- pi/po: Page In/ Out Rates (per second)
- fr: pages freed rate

iostat

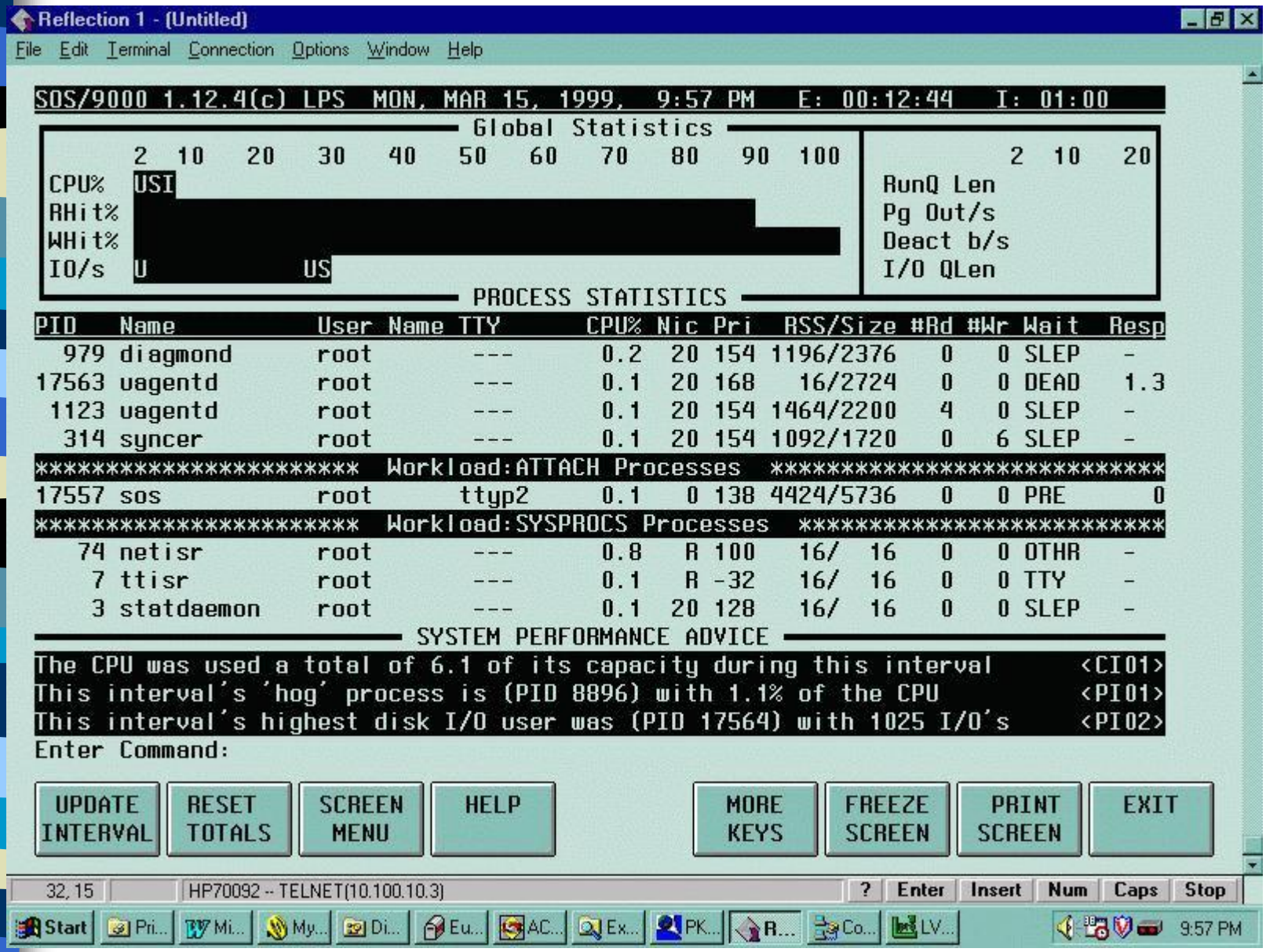
- Tin and tout-show char read and written
- CPU metrics - us, ni, sy, id
- bps - kilobytes per second, sps - seeks per second, msps - milliseconds per seek.

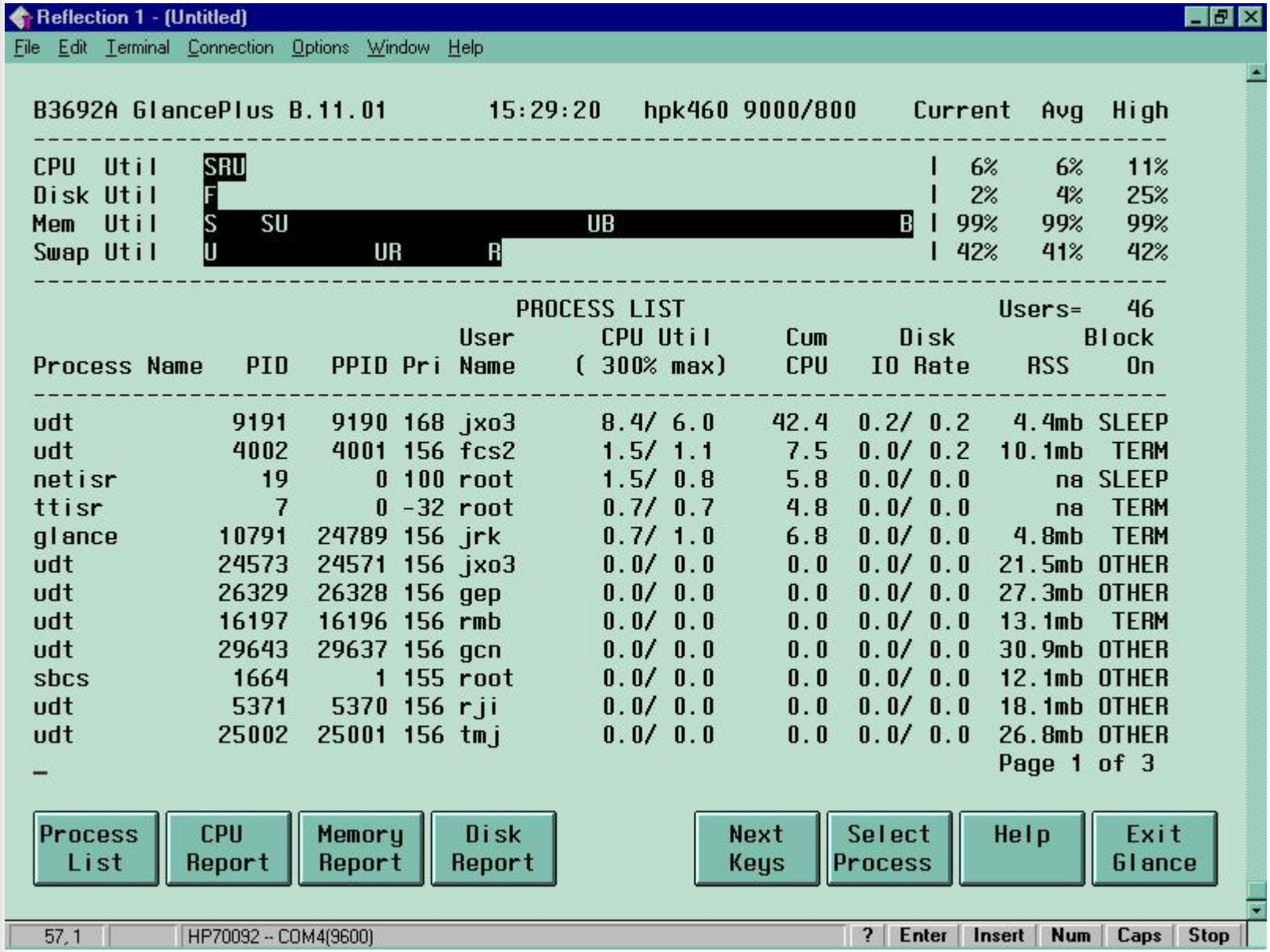
```
hpk460:/home/jrk$iostat -t
                 tty            cpu
               tin tout       us  ni  sy  id
                 1  219        2   0   2  96

  device    bps     sps    msps
  c2t6d0      0     0.0     1.0
  c2t5d0      0     0.0     1.0
  c1t0d2      0     0.0     1.0
 c2t10d0      0     0.0     1.0
 c2t11d0      0     0.0     1.0
```

Performance Monitoring tools

- Third party tools
  - HP products; Glance, GlancePlus, GPM, PerfRx, Perfview, Measureware.
  - Lund Tools: SOS Performance Advisor, Performance Gallery
  - TeamQuest, Sarcheck from Aurora Software.
  - BEST1 from BGS, EcoTools, BMC's Patrol, Athene, Viewpoint, Platinum.


Reflection 1 - (Untitled)
File Edit Terminal Connection Options Window Help
SOS/9000 1.12.4(c) LPS MON, MAR 15, 1999, 9:57 PM E: 00:12:44 I: 01:00
Global Statistics
2 10 20 30 40 50 60 70 80 90 100
CPU% USI
RHit%
WHit%
IO/s U US
2 10 20
RunQ Len
Pg Out/s
Deact b/s
I/O QLen
PROCESS STATISTICS
PID Name User Name TTY CPU% Nic Pri RSS/Size #Rd #Wr Wait Resp
979 diagmond root --- 0.2 20 154 1196/2376 0 0 SLEP -
17563 uagentd root --- 0.1 20 168 16/2724 0 0 DEAD 1.3
1123 uagentd root --- 0.1 20 154 1464/2200 4 0 SLEP -
314 syncer root --- 0.1 20 154 1092/1720 0 6 SLEP -
********** Workload:ATTACH Processes **********
17557 sos root ttyp2 0.1 0 138 4424/5736 0 0 PRE 0
********** Workload:SYSPROCS Processes **********
74 netisr root --- 0.8 R 100 16/ 16 0 0 OTHR -
7 ttisr root --- 0.1 R -32 16/ 16 0 0 TTY -
3 statdaemon root --- 0.1 20 128 16/ 16 0 0 SLEP -
SYSTEM PERFORMANCE ADVICE
The CPU was used a total of 6.1 of its capacity during this interval <CI01>
This interval's 'hog' process is (PID 8896) with 1.1% of the CPU <PI01>
This interval's highest disk I/O user was (PID 17564) with 1025 I/O's <PI02>
Enter Command:
UPDATE INTERVAL
RESET TOTALS
SCREEN MENU
HELP
MORE KEYS
FREEZE SCREEN
PRINT SCREEN
EXIT
32, 15
HP70092 -- TELNET(10.100.10.3)
? Enter Insert Num Caps Stop
Start
9:57 PM


Reflection 1 - (Untitled)
File Edit Terminal Connection Options Window Help
B3692A GlancePlus B.11.01 15:29:20 hpk460 9000/800 Current Avg High
CPU Util | 6% 6% 11%
Disk Util | 2% 4% 25%
Mem Util | 99% 99% 99%
Swap Util | 42% 41% 42%
PROCESS LIST Users= 46
Process Name PID PPID Pri User Name CPU Util (300% max) Cum CPU Disk IO Rate RSS Block On
udt 9191 9190 168 jxo3 8.4/ 6.0 42.4 0.2/ 0.2 4.4mb SLEEP
udt 4002 4001 156 fcs2 1.5/ 1.1 7.5 0.0/ 0.2 10.1mb TERM
netisr 19 0 100 root 1.5/ 0.8 5.8 0.0/ 0.0 na SLEEP
ttisr 7 0 -32 root 0.7/ 0.7 4.8 0.0/ 0.0 na TERM
glance 10791 24789 156 jrk 0.7/ 1.0 6.8 0.0/ 0.0 4.8mb TERM
udt 24573 24571 156 jxo3 0.0/ 0.0 0.0 0.0/ 0.0 21.5mb OTHER
udt 26329 26328 156 gep 0.0/ 0.0 0.0 0.0/ 0.0 27.3mb OTHER
udt 16197 16196 156 rmb 0.0/ 0.0 0.0 0.0/ 0.0 13.1mb TERM
udt 29643 29637 156 gcn 0.0/ 0.0 0.0 0.0/ 0.0 30.9mb OTHER
sbcs 1664 1 155 root 0.0/ 0.0 0.0 0.0/ 0.0 12.1mb OTHER
udt 5371 5370 156 rji 0.0/ 0.0 0.0 0.0/ 0.0 18.1mb OTHER
udt 25002 25001 156 tmj 0.0/ 0.0 0.0 0.0/ 0.0 26.8mb OTHER
Page 1 of 3
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
57,1
HP70092 -- COM4(9600)
? Enter Insert Num Caps Stop

Third Party Tools -- Advantages

- Integrated
- Many useful screens
- User friendly
- Function key or letter command driven
- Contain help entries, interpretations & alarms
- Easier to use for reports and trends

Analysis Points

- Best with long-term evaluation
- Require “rules of Thumb”
- Move step-by-step
- Change one thing at a time to find out what is most effective

Rules of Thumb

- Commonly held theorems about levels of acceptable usage.
  - CPU - total busy above 85 % is not good
  - CPU Queue Length of 5 is going to begin to show poor performance, 15 and above is very bad.
  - Memory - buffer cache read hit rate and write hit of less than 90 % is not good.

Rules of Thumb (continued)

- Capture Ratio (this is a ratio of the user processing / by system activity or (User + Real + Nice)/(System + Interrupt + Context Switch) = Capture ratio)). Should be greater than 3 and will definitely reflect a problem when 1.0.
- Real processing, System, Interrupt, Context Switching should not exceed 10 % as individual measures.

Rules of Thumb (continued)

- Memory % used should not exceed 80 to 90 %.
- Virtual % used should not exceed 50 to 80 %.
- Disk I/O Queue length should not exceed 1.0.
- Total reads & writes per drive should not exceed 50 to 60.

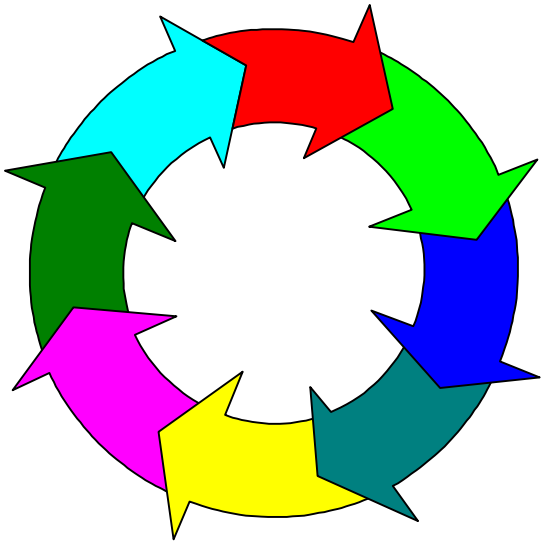
Rules of Thumb (continued)

- Page out rate of 10 per second is not good and an indicator of memory shortage.
- Deactivation rate of 5 processes is an indicator of memory issues.
- Network - watch errors and collisions

Step-by-step Analysis

- Begin Macro to Micro (CPU, Memory,disk to database engine and application)

•Remember - It is not a once and you are done! It is more of a cycle.



Step by Step

- Examine CPU, than memory, than disk, network, program and database.
- Realize the prevalance of disk issues.
- Collect data.
- Get training, books, etc.
- Check system configuration.
- Kernel parameters

Kernel Tuning

- Larger issue, here are some recommendations:
  - bufpages 0
  - create_fastlinks 1
  - dbc_max_pct 25
  - fs_async 0
  - nflocks = (nfile/4)

Database Engine issues

- Very significant area.
- Data loads and unloads, re-indexing, etc.
- Be sure you have the appropriate indexes.
- Understand your I/O.
- Look at vendor recommendations

Performance Baseline

- A picture or an understanding of what performance was like when response time was defined as “good” or acceptable.
- Helpful in the identification of performance problems.

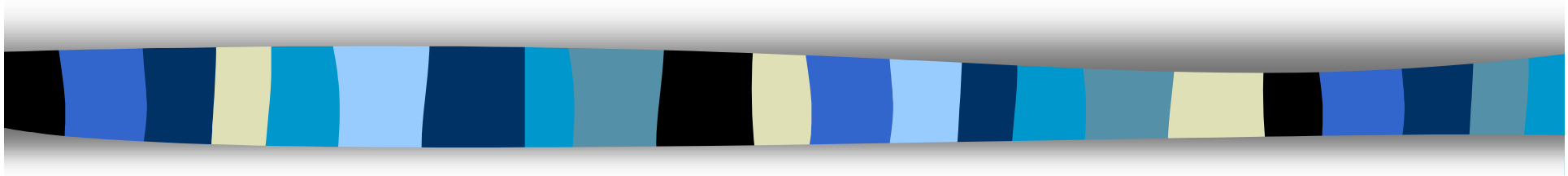
Conclusion

- Maximize investment
- Avoid rushed “bad” decisions
- Avoid the unforeseen
- Importance of performance, this area needs attention!

Performance Problems

- Past, Present, and Future
- Crises, problem solving, learning & planning
- Do the following:
  - observe system usage
  - understand their meaning.
  - “rules of thumb”
  - Outline a plan
  - Make changes

The End



Thanks for coming!

Signup with HP-UX performance email group: hpuxperform@interex.org

Informative URL://http://docs.hp.com/hpux/content/KCparams.OverviewAll.html