



# **Performance and Workload Management**



# Unit Objectives

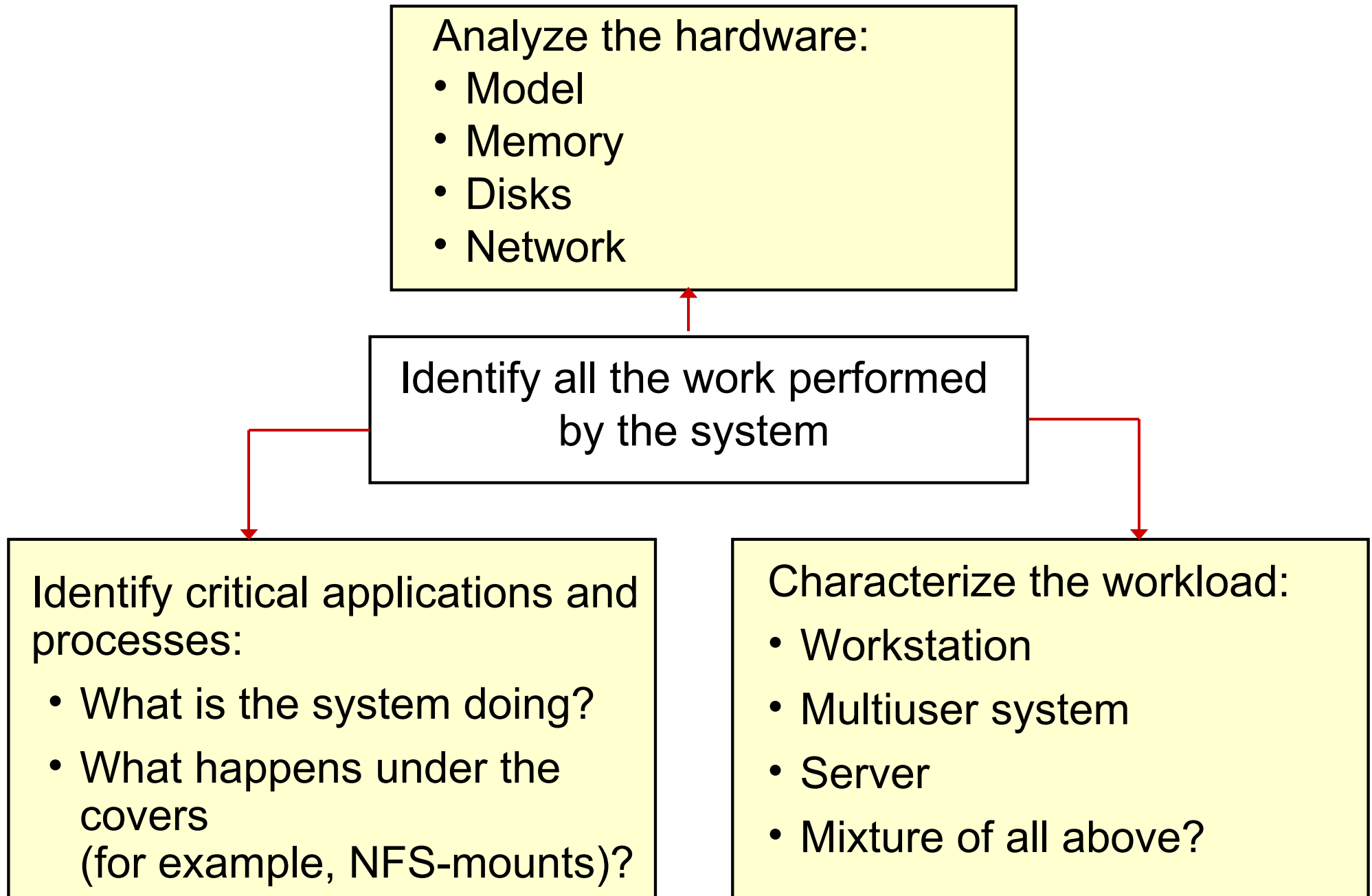
---

After completing this unit, you should be able to:

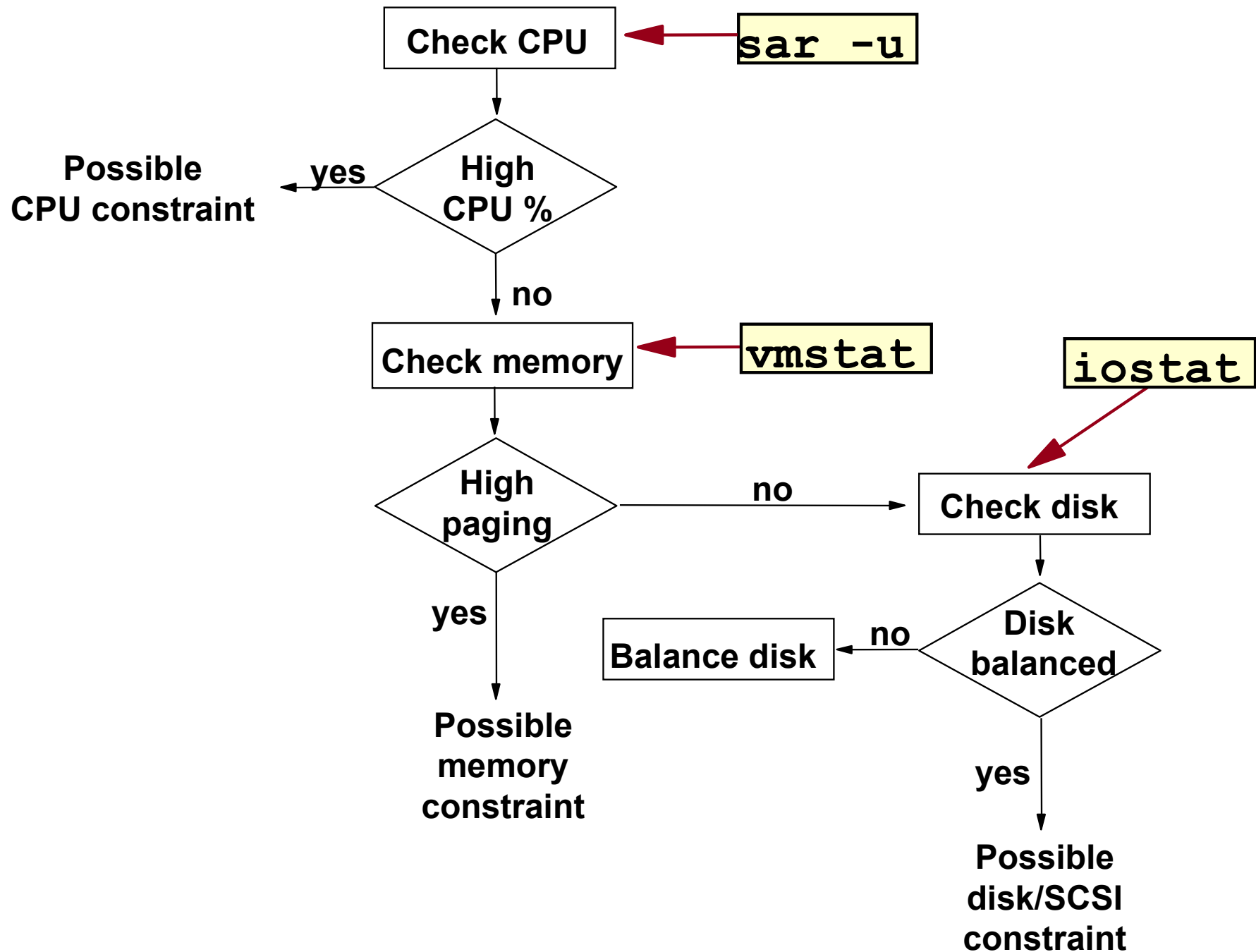
- Provide basic performance concepts
- Provide basic performance analysis
- Manage the workload on a system
- Use the Performance Diagnostic Tool (PDT)

# Understand the Workload

---



# Basic Performance Analysis



# Identify CPU-Intensive Programs: `ps aux`

# `ps aux`

USER	PID	%CPU	%MEM	...	STIME	TIME	COMMAND
root	516	98.2	0.0	...	13:00:00	1329:38	wait
johnp	7570	1.2	1.0	...	17:48:32	0:01	-ksh
root	1032	0.8	0.0	...	15:13:47	78:37	kproc
root	1	0.1	1.0	...	15:13:50	13:59	/etc/init

Percentage of time  
the process has  
used the CPU

Percentage of real  
memory

Total Execution  
Time

# Identify High Priority Processes: `ps -elf`

# `ps -elf`

	F	S	UID	PID	PPID	C	PRI	NI	...	TIME	CMD
200003	A	root	1	0	0	60	20	...	0:04	/etc/init	
240001	A	root	69718	1	0	60	20	...	1:16	/usr/sbin/syncd 60	
200001	A	root	323586	188424	24	72	20	...	0:00	ps -elf	

Priority of  
the process

Nice value

- The smaller the PRI value, the higher the priority of the process. The average process runs a priority around 60.
- The NI value is used to adjust the process priority. The higher the nice value is, the lower the priority of the process.

# Monitoring CPU Usage: `sar -u`

Interval

Number

```
# sar -u 60 30
```

```
AIX www 3 5 000400B24C00 08/09/05
```

```
System configuration: lcpu=2
```

08:24:10	%usr	%sys	%wio	%idle
08:25:10	48	52	0	0
08:26:10	63	37	0	0
08:27:10	59	41	0	0
...				
Average	<b>57</b>	<b>43</b>	0	0

A system may be CPU bound, if:  
 $\%usr + \%sys > 80\%$

# AIX Tools: tprof

```
# tprof -x sleep 60
# more sleep.prof
```

Process	Freq	Total	Kernel	User	Shared	Other
=====	=====	=====	=====	=====	=====	=====
./cpuprog	5	99.56	92.86	3.05	3.64	0.00
/usr/bin/tprof	2	0.41	0.01	0.01	0.39	0.00
/usr/sbin/syncd	4	0.02	0.02	0.00	0.00	0.00
gil	2	0.01	0.01	0.00	0.00	0.00
/usr/bin/sh	1	0.00	0.00	0.00	0.00	0.00
/usr/bin/trcstop	1	0.00	0.00	0.00	0.00	0.00
=====	=====	=====	=====	=====	=====	=====
Total	15	100.00	92.91	3.06	4.03	0.00

Process	PID	TID	Total	Kernel	User	Shared	Other
=====	=====	=====	=====	=====	=====	=====	=====
./cpuprog	184562	594051	20.00	18.72	0.63	0.66	0.00
./cpuprog	262220	606411	19.96	18.64	0.58	0.74	0.00
./cpuprog	168034	463079	19.89	18.57	0.61	0.71	0.00
./cpuprog	254176	598123	19.87	18.51	0.61	0.74	0.00
./cpuprog	282830	618611	19.83	18.43	0.61	0.79	0.00
/usr/bin/tprof	270508	602195	0.40	0.01	0.01	0.39	0.00
/usr/sbin/syncd	73808	163995	0.01	0.01	0.00	0.00	0.00
/usr/bin/trcstop	196712	638993	0.00	0.00	0.00	0.00	0.00
/usr/bin/sh	196710	638991	0.00	0.00	0.00	0.00	0.00
gil	49176	61471	0.00	0.00	0.00	0.00	0.00
...							
=====	=====	=====	=====	=====	=====	=====	=====
Total			100.00	92.91	3.06	4.03	0.00

Total Samples = 24316

Total Elapsed Time = 121.59s

# Monitoring Memory Usage: `vmstat`

Summary report every 5 seconds

```
# vmstat 5
```

System Configuration: `lcpu=2 mem=512MB`

kthr		memory			page					...	cpu			
r	b	avm	fre	re	pi	po	fr	sr	cy	...	us	sy	id	wa
0	0	8793	81	0	0	0	1	7	0		1	2	95	2
0	0	9192	66	0	0	16	81	167	0		1	6	77	16
0	0	9693	69	0	0	53	95	216	0		1	4	63	33
0	0	10194	64	0	21	0	0	0	0		20	5	42	33
0	0	4794	5821	0	24	0	0	0	0		5	8	41	46

`pi, po:`

- Paging space page ins and outs
- If any paging space I/O is taking place, the workload is approaching the system's memory limit

`wa:`

- I/O wait percentage of CPU
- If non-zero, a significant amount of time is being spent waiting on file I/O

# AIX Tools: svmon

## Global report

# **svmon -G**

	size	inuse	free	pin	virtual
memory	32744	20478	12266	2760	11841
pg space	65536	294			
		work	pers	clnt	lpage
pin		2768	0	0	0
in use		13724	6754	0	0

Sizes are in # of  
4K frames

## Top 3 users of memory

# **svmon -Pt 3**

Pid	Command	Inuse	Pin	Pgsp	Virtual	64-bit	Mthrd	Lpage
14624	java	6739	1147	425	4288	N	Y	N
...								
9292	httpd	6307	1154	205	3585	N	Y	N
...								
3596	X	6035	1147	1069	4252	N	N	N
...								

\* output has been modified

# Monitoring Disk I/O: iostat

```
# iostat 10 2
```

```
System configuration: lcpu=2 drives=3 ent=0.30 paths=4 vdisks=1
```

tty:	tin	tout	avg-cpu:	%user	%sys	%idle	%iowait	phycs	%entc
	0.1	110.7		7.0	59.4	0.0	<b>33.7</b>	<b>0.0</b>	<b>1.4</b>

Disks:	% <b>tm_act</b>	Kbps	tps	Kb_read	Kb_wrtn
--------	-----------------	------	-----	---------	---------

hdisk0	<b>77.9</b>	115.7	28.7	456	8
hdisk1	0.0	0.0	0.0	0	0
cd0	0.0	0.0	0.0	0	0

tty:	tin	tout	avg-cpu:	%user	%sys	%idle	%iowait	phycs	%entc
	0.1	96.3		6.5	58.0	0.0	<b>35.5</b>	<b>0.0</b>	<b>1.3</b>

Disks:	% <b>tm_act</b>	Kbps	tps	Kb_read	Kb_wrtn
--------	-----------------	------	-----	---------	---------

hdisk0	<b>79.8</b>	120.1	28.7	485	9
hdisk1	0.0	0.0	0.0	0	0
cd0	0.0	0.0	0.0	0	0

# AIX Tools: filemon

```
# filemon -o fmout
```

Starts monitoring disk activity

```
# trcstop  
# more fmout
```

Stops monitoring and creates report

## Most Active Logical Volumes

util	#rblk	#wblk	KB/s	volume	description
0.03	3368	888	26.5	/dev/hd2	/usr
0.02	0	1584	9.9	/dev/hd8	jfs2log
0.02	56	928	6.1	/dev/hd4	/

## Most Active Physical Volumes

util	#rblk	#wblk	KB/s	volume	description
0.10	24611	12506	231.4	/dev/hdisk0	Virtual SCSI Disk Drive
0.02	56	8418	52.8	/dev/hdisk1	N/A

# topas

## # topas

```
Topas Monitor for host:      kca81
Mon Aug  9 11:48:35 2005    Interval:  2

EVENTS/QUEUES      FILE/TTY
Cswitch            370  Readch            11800
Syscall            461  Writtech             95
Reads              18   Rawin                0
Writes              0   Ttyout                0
Forks               0   Igets                 0
Execs               0   Namei                 1
Runqueue           0.0  Dirblk                0
Waitqueue           0.0

Kernel    0.1   |                               |
User      0.0   |                               |
Wait      0.0   |                               |
Idle     99.8   |#####|
Physc =   0.00                               %Entc=   1.5

Network  KBPS   I-Pack  O-Pack   KB-In  KB-Out
en0      0.1     0.4     0.4     0.0    0.1
lo0      0.0     0.0     0.0     0.0    0.0

Disk      Busy%   KBPS     TPS  KB-Read  KB-Writ
hdisk0    0.0     0.0     0.0    0.0     0.0
hdisk1    0.0     0.0     0.0    0.0     0.0

PAGING      MEMORY
Faults       1   Real,MB      4095
Steals        0   % Comp        15.4
PgspIn        0   % Noncomp      9.3
PgspOut        0   % Client       1.8
PageIn         0
PageOut         0   PAGING SPACE
Sios           0   Size,MB       3744
                % Used        0.6
                % Free        99.3

Name          PID CPU%  PgSp  Owner
topas         18694 0.1   1.4  root
rmcd          10594 0.0   2.0  root
nfsd          15238 0.0   0.0  root
syncd         3482 0.0   1.3  root
gil           2580 0.0   0.0  root

NFS (calls/sec)
ClientV2       0   WPAR Activ   0
ServerV2       0   WPAR Total   0
ClientV2       0   Press:
ServerV3       0   "h" for help
ClientV3       0   "q" for quit
```

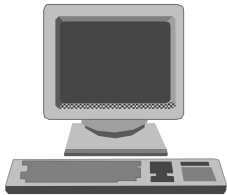
CPU  
info

iostat  
info

vmstat  
info

# There Is Always a Next Bottleneck!

---



```
# iostat 10 60
```

Our system is I/O bound.  
Let's buy faster disks!



```
# vmstat 5
```

Our system is now  
memory bound! Let's buy  
more memory!!!



```
# sar -u 60 60
```

Oh no! The CPU is  
completely overloaded!

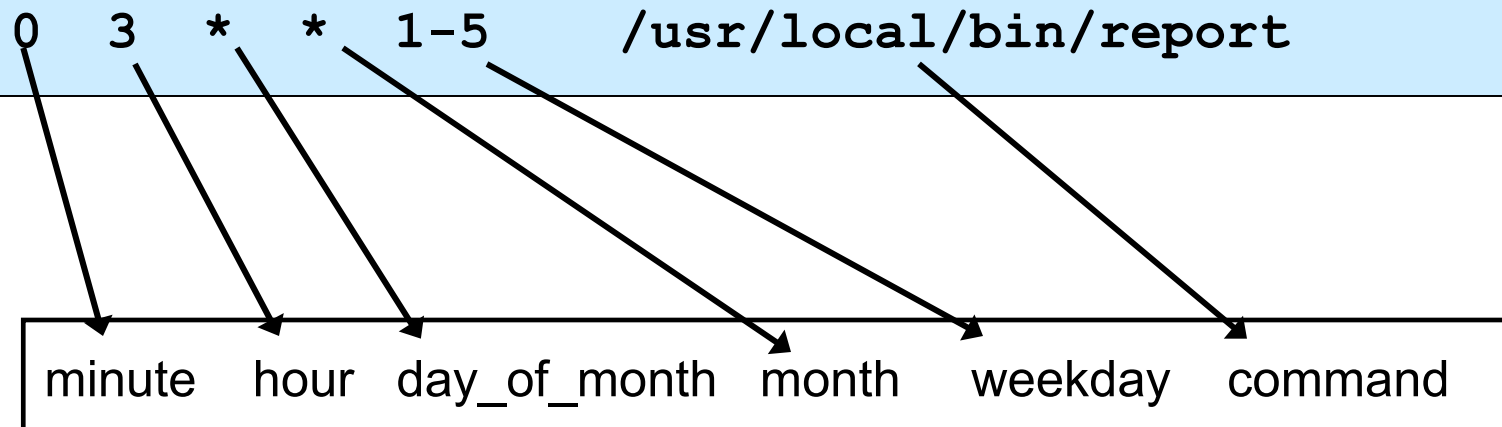
# Workload Management Techniques (1 of 3)

Run programs at a specific time

```
# echo "/usr/local/bin/report" | at 0300
# echo "/usr/bin/cleanup" | at 1100 friday
```

```
# crontab -e
```

```
0 3 * * 1-5 /usr/local/bin/report
```



# Workload Management Techniques (2 of 3)

## Sequential execution of programs

```
# vi /etc/qconfig
```

```
ksh:
```

```
    device = kshdev  
    discipline = fcfs
```

```
kshdev:
```

```
    backend = /usr/bin/ksh
```

```
# qadm -D ksh
```

Queue is down

```
# qprt -P ksh report1
```

```
# qprt -P ksh report2
```

```
# qprt -P ksh report3
```

Jobs will be queued

```
# qadm -U ksh
```

Queue is up:  
Jobs will be executed  
sequentially

# Workload Management Techniques (3 of 3)

Run programs at a reduced priority

```
# nice -n 15 backup_all &
```

```
# ps -el
```

F	S	UID	PID	PPID	C	PRI	NI	...	TIME	CMD
240001	A	0	3860	2820	30	<b>90</b>	<b>35</b>	...	0:01	backup_all

Very low  
priority

Nice value:  
20+15

```
# renice -n -10 3860
```

```
# ps -el
```

F	S	UID	PID	PPID	C	PRI	NI	...	TIME	CMD
240001	A	0	3860	2820	26	<b>78</b>	<b>25</b>	...	0:02	backup_all

# Simultaneous Multi-Threading (SMT)

---

- Each chip appears as a two-way SMP to software:
  - Appear as 2 logical CPUs
  - Performance tools may show number of logical CPUs
- Processor resources optimized for enhanced SMT performance:
  - May result in a 25-40% boost and even more
- Benefits vary based on workload
- To enable:  
`smtctl [ -m off | on [ -w boot | now ] ]`

# Tool Enhancements for Micro-Partitioning

---

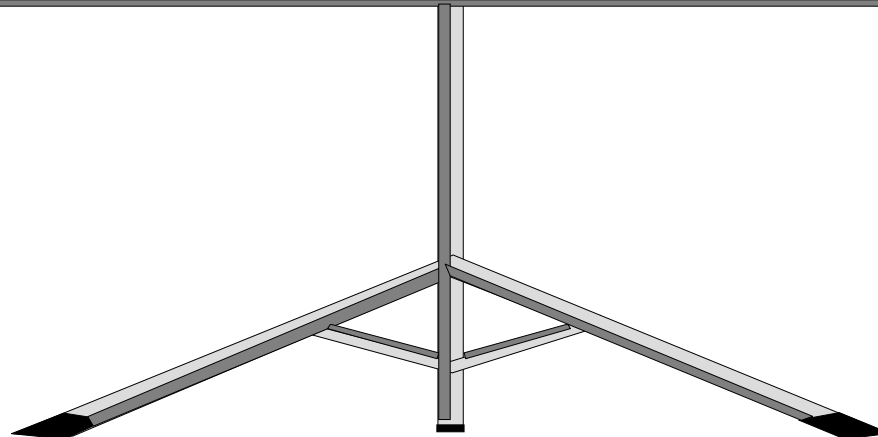
- Added two new values to the default **topas** screen
  - **Phy**sc and %**Ent**c
- The **vmstat** command has two new metrics:
  - **pc** and **ec**
- The **iostat** command has two new metrics:
  - %**physc** and %**entc**
- The **sar** command has two new metrics:
  - **physc**
  - %**entc**

# Exercise 12: Basic Performance Commands

---



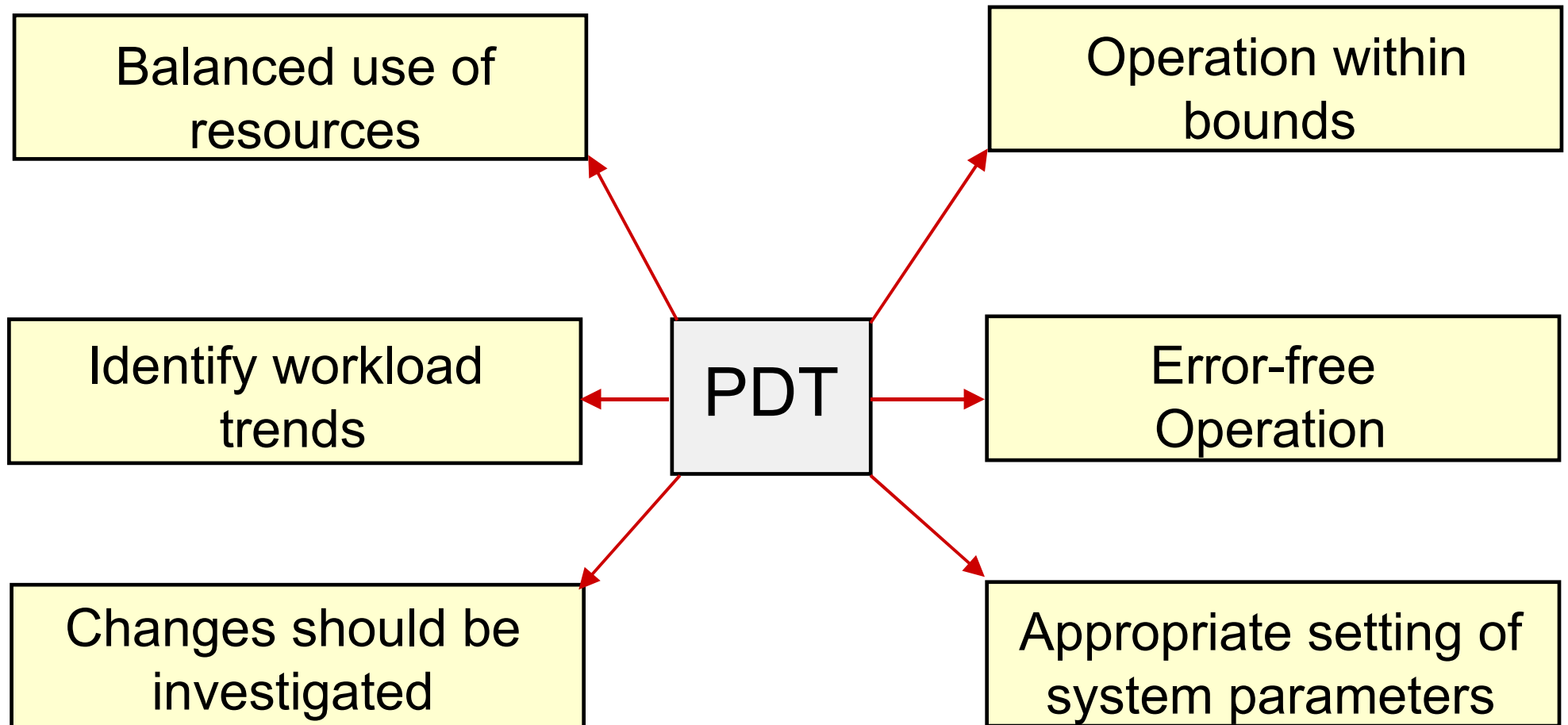
- Working with **ps**, **nice**, and **renice**
- Basic performance analysis
- Working with a Korn shell job queue



# Performance Diagnostic Tool (PDT)

---

PDT assesses the current state of a system and tracks changes in workload and performance.



# Enabling PDT

---

```
# /usr/sbin/perf/diag_tool/pdt_config
```

```
-----PDT customization menu-----
```

- show current PDT report recipient and severity level
- modify/enable PDT reporting
- disable PDT reporting
- **modify/enable PDT collection**
- disable PDT collection
- de-install PDT
- exit pdt\_config

```
Please enter a number: 4
```

# cron Control of PDT Components

---

```
# cat /var/spool/cron/crontabs/adm
```

```
0 9 * * 1-5 /usr/sbin/perf/diag_tool/Driver_ daily
```

Collect system data, each workday at 9:00 A.M.

```
0 10 * * 1-5 /usr/sbin/perf/diag_tool/Driver_ daily2
```

Create a report, each workday at 10:00 A.M.

```
0 21 * * 6 /usr/sbin/perf/diag_tool/Driver_ offweekly
```

Clean up old data, each Saturday at 9:00 P.M.

# PDT Files

## Collection

Driver\_daily  
/var/perf/cfg/diag\_tool/.collection.control

## Retention

Driver\_offweekly  
/var/perf/cfg/diag\_tool/.retention.control

## Reporting

Driver\_daily2  
/var/perf/cfg/diag\_tool/.reporting.control

/var/perf/tmp/.SM

→ /var/perf/tmp/.SM.last

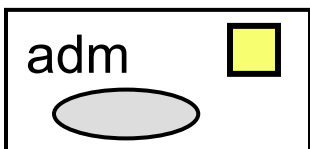
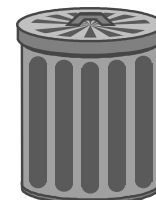
/var/perf/tmp/PDT\_REPORT

Next Day

/var/perf/tmp/PDT\_REPORT.last

35 days  
.retention.list

/var/perf/tmp/.SM.discards



# Customizing PDT: Changing Thresholds

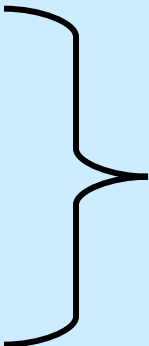
---

```
# vi    /var/perf/cfg/diag_tool/.thresholds

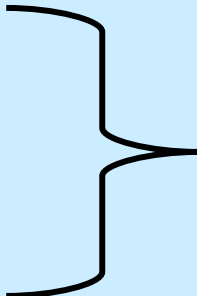
DISK_STORAGE_BALANCE 800
PAGING_SPACE_BALANCE 4
NUMBER_OF_BALANCE 1
MIN_UTIL 3
FS_UTIL_LIMIT 90
MEMORY_FACTOR .9
TREND_THRESHOLD .01
EVENT_HORIZON 30
```

# Customizing PDT: Specific Monitors

```
# vi /var/perf/cfg/diag_tool/.files
```

<b>/var/adm/wtmp</b>		<b>Files and directories to monitor</b>
<b>/var/spool/qdaemon/</b>		
<b>/var/adm/ras/</b>		
<b>/tmp/</b>		

```
# vi /var/perf/cfg/diag_tool/.nodes
```

<b>pluto</b>		<b>Systems to monitor</b>
<b>neptun</b>		
<b>mars</b>		

# PDT Report Example (Part 1)

---

## **Performance Diagnostic Facility 1.0**

Report printed: Sun Aug 21 20:53:01 2005

Host name: master

Range of analysis included measurements  
from: Hour 20 on Sunday, August 21st, 2005  
to: Hour 20 on Sunday, August 21st, 2005

## **Alerts**

### **I/O CONFIGURATION**

- Note: volume hdisk2 has 480 MB available for allocation while volume hdisk1 has 0 MB available

### **PAGING CONFIGURATION**

- Physical Volume hdisk1 (type:SCSI) has no paging space defined

### **I/O BALANCE**

- Physical volume hdisk0 is significantly busier than others  
volume hdisk0, mean util. = 11.75  
volume hdisk1, mean util. = 0.00

### **NETWORK**

- Host sys1 appears to be unreachable

# PDT Report Example (Part 2)

---

## Upward Trends

### FILES

- File (or directory) /var/adm/ras/ SIZE is increasing  
now, 364 KB and increasing an avg. of 5282 bytes/day

### FILE SYSTEMS

- File system lv01(/fs3) is growing  
now, 29.00% full, and growing an avg. of 0.30%/day  
At this rate lv01 will be full in about 45 days

### ERRORS

- Hardware ERRORS; time to next error is 0.982 days

## System Health

### SYSTEM HEALTH

- Current process state breakdown:  
2.10 [0.5%]: waiting for the CPU  
89.30 [22.4%]: sleeping  
306.60 [77.0%]: zombie  
398.00 = TOTAL

## Summary

This is a severity level 1 report

No further details available at severity level >1

# Checkpoint

1. What commands can be executed to identify CPU-intensive programs?
  - 
  -
- What command can be executed to start processes with a lower priority? \_\_\_\_\_
5. What command can you use to check paging I/O? \_\_\_\_\_
7. True or False? The higher the PRI value, the higher the priority of a process.

# Checkpoint Solutions

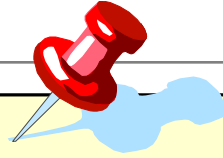
1. What commands can be executed to identify CPU-intensive programs?
  - `ps aux`
  - `top`
3. What command can be executed to start processes with a lower priority? `nice`
5. What command can you use to check paging I/O? `vmstat`
- True or False? The higher the PRI value, the higher the priority of a process.

# Exercise 13: Performance Diagnostic Tool

---

- 
- Use the Performance Diagnostic Tool to:
    - Capture data
    - Create reports

# Unit Summary



- The following commands can be used to identify potential bottlenecks in the system:
  - `ps`
  - `sar`
  - `vmstat`
  - `iostat`
- If you cannot fix a performance problem, manage your workload through other means (`at`, `crontab`, `nice`, `renice`).
- Use the Performance Diagnostic tool (PDT) to assess and control your systems performance.