A21

Topas Monitoring and Recording Tool

Luc Smolders

**IBM SYSTEM p, AIX 5L
and LINUX TECHNICAL
UNIVERSITY
Sept 11 - 15, 2006**

Las Vegas, NV

charts to be available at:
http://www.ibm.com/developerworks/blogs/page/Systemptechuniv

Agenda

- Introduction
  - history
  - technology
- Basic real-time monitoring functions
  - main screen
  - full-screen process monitoring
- Support for WLM
- Detailed Disk monitoring
- Support for SMT and micro-partitions
- CEC monitoring
- Automatic performance metrics recording
  - daily, WLE, CEC
  - report generation and recorded data exports
- PTX affinity
- Future plans

topas - introduction

- History
  - started life as a PTX API sample program (lchmon)
  - enhanced into a top-like monitoring tool
- Technology
  - curses application
  - data collection uses PTX APIs
    - ► SPMI for local data
    - ► RSI for remore data (not in initial release)
- Enhancements
  - support for WLM
    - ► main panel
    - ► dedicated panel
  - support for SMT/SPLPAR
    - ► new way (PURR based) to calculate cpu utilization
    - ► lparstat like panel
  - support for CEC-wide monitoring
    - ► uses xmtopas daemons and RSI APIs
    - ► new panel
  - support for detailed disk statistics
    - ► dedicated panel
  - recordings
    - ► initial release for WLE data collection, local data only
    - ► latest release includes CEC data
    - ► topasout text reports

topas - available options

- -h
  - help information

- -i interval
  - monitoring interval in seconds
- -m
  - use monochrome mode - no colors

- -d number
  - number of monitored disks
- -n number
  - number of monitored network interfaces
- -p number
  - number of monitored processes
- -w number
  - number of monitored WLM classes
- -c number
  - number of monitored CPUs

- -P
  - show full-screen Process Display
- -U username
  - show username owned processes with -P

- -L
  - show full-screen Logical Partition display
- -D
  - show full-screen Disk display
- -W
  - show full-screen WLM display

- -R
  - records Cross-LPAR (CEC) metrics
- -C
  - show full-screen Cross-LPAR (CEC) display
- -I interval
  - remote monitoring interval in seconds
- -o
  - sets CEC panel configuration parameters
  - syntax: -o parameter=value

topas - main screen

```
Topas Monitor for host:        lambic
Sat Sep 23 16:45:06 2000    Interval:  2

Kernel    1.5    |                              |
User      1.0    |                              |
Wait      0.0    |                              |
Idle     97.4    |########################### |

Interf      KBPS     I-Pack    O-Pack     KB-In   KB-Out
tr0          2.5        7.0       8.5       0.4      2.1
lo0          1.2        2.0       2.0       0.6      0.6

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

Name            Pid CPU% PgSp  Owner
java          10064  1.0  3.8  root
gil            1032  0.5  0.3  root
topas         11644  0.5  0.4  root
aixterm       11100  0.0  0.8  steve
ksh            4136  0.0  0.5  root
syncd          4406  0.0  0.3  root
ksh           10424  0.0  0.2  root
xmservd       10850  0.0  0.7  root
ksh            9836  0.0  0.5  steve
netm            774  0.0  0.2  root
init              1  0.0  0.7  root
cron           5244  0.0  0.5  root
ksh           10674  0.0  0.5  steve
snmpd          6966  0.0  0.9  root
```

```
EVENTS/QUEUES        FILE/TTY
Cswitch        221   Readch       3873
Syscall        864   Writech      1870
Reads           11   Rawin           0
Writes           9   Ttyout        180
Forks            0   Igets           0
Execs            0   Namei           6
Runqueue       0.0   Dirblk          0
Waitqueue      0.0

PAGING               MEMORY
Faults          38   Real,MB       255
Steals           0   % Comp       28.0
PgspIn           0   % Noncomp    11.0
PgspOut          0   % Client      0.0
PageIn           0
PageOut          0   PAGING SPACE
Sios             0   Size,MB       512
                     % Used        0.2
NFS (calls/sec)      % Free        99.7
ServerV2         0
ClientV2         0      Press:
ServerV3         0      "h" for help
ClientV3         0      "q" to quit
```

fixed areas

sortable by any column
using tabs and arrows

customizable areas

topas – Events/Queues and File/tty metrics

- Events/Queues
  - Cswitch
    - number of context switches per second
  - Syscalls
    - total number of system calls per second
  - Reads
    - number of read system calls per second
  - Writes
    - number of write system calls per second
  - Forks
    - number of fork system calls per second
  - Execs
    - number of exec system calls per second
  - Runqueue
    - average number of threads that were ready to run but were waiting for a processor to become available
  - Waitqueue
    - average number of threads that were waiting for paging to complete

- File/tty
  - Readch
    - number of bytes read per second through the read system call
  - Writech
    - number of bytes written per second through the write system call
  - Rawin
    - number of raw bytes read per second from TTYs
  - Ttyout
    - number of bytes written to TTYs per second
  - Igets
    - number of calls per second to the inode lookup routines
  - Namei
    - number of calls per second to the pathname lookup routines
  - Dirblk
    - number of directory blocks scanned per second by the directory search routine

topas - Paging, Memory and Paging Space metrics

- Paging
  - Faults
    - total number of page faults per second
      - includes page faults that do not cause paging activity
  - Steals
    - physical memory 4K frames stolen per second by the virtual memory manager(LRU daemon)
  - PgspIn
    - number of 4K pages read from paging space per second
  - PgspOut
    - number of 4K pages written to paging space per second
  - PageIn
    - number of 4K pages read per second
      - includes paging activity associated with reading from file systems
      - subtract PgspIn from this value to get the number of 4K pages read from file systems per second
  - PageOut
    - number of 4K pages written per second
      - includes paging activity associated with writing to file systems
      - subtract PgspOut from this value to get the number of 4K pages written to file systems per second
  - Sios
    - number of I/O requests per second issued by the virtual memory manager

- Memory
  - Real,MB
    - size of real memory in megabytes.
  - % Comp
    - percentage of real memory currently allocated to computational page frames
      - computational page frames are generally those that are backed by paging space.
  - % Noncomp
    - percentage of real memory currently allocated to non-computational frames
      - non-computational page frames are generally those that are backed by file space, either data files, executable files, or shared library files.
  - % Client
    - percentage of real memory currently allocated to cache remotely mounted files or JFS2 files

- Paging Space
  - Size,MB
    - sum of all paging spaces on the system, in megabytes.
  - % Used
    - percentage of total paging space currently in use.
  - % Free
    - percentage of total paging space currently free.

topas - main screen command

- One-character commands
  - a - Show all the variable subsections being monitored
    - ► pressing the the 'a' key always returns topas to the main inital display
  - c - Pressing the 'c' key repeatedly toggles the CPU subsection between
    - ► cumulative report
    - ► off
    - ► list of busiest CPUs on the system
  - d - Pressing the 'd' key repeatedly toggles the disk subsection between
    - ► busiest disks list
    - ► off
    - ► total disk activity for the system.
  - n - Pressing the 'n' key repeatedly toggles the network interfaces subsection between
    - ► busiest interfaces list
    - ► off
    - ► total network activity for the system
  - p - Pressing the 'p' key toggles the hot processes subsection on and off
  - w - Pressing the 'w' key toggles the WorkLoad Management(WLM) classes subsection on and off

  - q - Quit the program
  - r - Refresh the screen immediately

  - P - Toggle to the Full Screen Process Display
  - W - Toggle to the Full Screen WLM Class Display
  - L - Toggle to the Logical Partition/CPU Display
  - D - Toggle to the Disk Display
  - C - Toggle to the Cross-LPAR (CEC) Panel

- The network, disk, cpu, and WLM sections may be sorted using arrows and tabs. A highlighted column header shows sorted-by

topas - main screen(2)

```
Topas Monitor for host:      lambic          EVENTS/QUEUES          FILE/TTY
Wed Jan 22 12:13:44 2003    Interval:  2     Cswitch          46   Readch           0
                                             Syscall          76   Writech         31
CPU       User%    Kern%    Wait%    Idle%   Reads             0   Rawin            0
cpu0       0.5      0.5      0.0     99.0    Writes            0   Ttyout           0
                                             Forks             0   Igets            0
Network  KBPS    I-Pack  O-Pack    KB-In  KB-Out   Execs      0   Namei            0
en0       0.0      0.0     0.0      0.0     0.0    Runqueue    0.0  Dirblk          0
tr0       0.0      0.0     0.0      0.0     0.0    Waitqueue   0.0
lo0       0.0      0.0     0.0      0.0     0.0
                                             PAGING                MEMORY
Disk     Busy%     KBPS      TPS KB-Read KB-Writ   Faults     0   Real,MB        255
Total     0.0       0.0      0.0     0.0     0.0   Steals      0   % Comp        32.5
                                             PgspIn            0   % Noncomp     14.3
Name          PID CPU% PgSp Owner           PgspOut           0   % Client       0.5
topas       13704  0.5  0.9 root            PageIn            0
rpc.lockd    8520  0.5  0.0 root            PageOut           0   PAGING SPACE
aixmibd      7482  0.0  0.8 root            Sios              0   Size,MB        512
snmpmibd     7232  0.0  0.5 root                                  % Used         1.2
gil          1806  0.0  0.0 root            NFS (calls/sec)       % Free        98.7
syncd        4182  0.0  0.6 root            ServerV2          0
netm         1548  0.0  0.0 root            ClientV2          0   Press:
init            1  0.0  0.7 root            ServerV3          0   "h" for help
cron         8278  0.0  0.4 root            ClientV3          0   "q" to quit
cron         5244  0.0  0.5 root
ksh         10674  0.0  0.5 steve
snmpd        6966  0.0  0.9 root
```

- Display after hitting "ccdd"
  - toggling cpu section to list of cpus
  - toggling disk section to one line total summary

topas - full screen process view

```
Topas Monitor for host:      lambic       Interval:    2     Wed Jan 22 12:10:44 2003

                                DATA  TEXT  PAGE                PGFAULTS
USER        PID   PPID PRI NI   RES   RES  SPACE    TIME CPU%  I/O  OTH COMMAND
root          1      0  60 20   202     9    202    0:07  0.0    0    0 init
root        774      0  17 41     4     0      4    0:00  0.0    0    0 reaper
root       1032      0  16 41     3     0      3    0:00  0.0    0    0 lrud
root       1290      0  60 41     4     0      4    0:04  0.0    0    0 xmgc
root       1548      0  36 41     4     0      4    0:07  0.0    0    0 netm
root       1806      0  37 41    16     0     16    7:09  0.0    0    0 gil
root       2064      0  16 41     4     0      4    0:00  0.0    0    0 wlmsched
root       2718      1  60 20   183    22    183    0:00  0.0    0    0 errdemon
root       2898   6984  60 20   131    20    131    0:03  0.0    0    0 hostmibd
root       3372      0  60 20     4     0      4    0:00  0.0    0    0 lvmbb
root       3628      0  50 41     4     0      4    0:00  0.0    0    0 jfsz
root       4182      1  60 20   155     2    155    1:41  0.0    0    0 syncd
root       4544   6984  60 20   106    11    106    0:00  0.0    0    0 inetd
root       4628      1  60 20     4     0      4    0:00  0.0    0    0 random
daemon     4950   6984  60 20   221    20    221    0:00  0.0    0    0 rpc.statd
root       5234   6984  60 20   244   105    244    0:03  0.0    0    0 sendmail
root       5472   6984  60 20    92     8     92    0:00  0.0    0    0 syslogd
root       5702   6984  60 20   133    12    133    0:00  0.0    0    0 portmap
root       5936   6984  60 20   271    39    271    0:00  0.0    0    0 snmpdv3ne
root       6450      0  60 20     4     0      4    0:00  0.0    0    0 rtcmd
```

- Output from topas -P or "P" command
- Sorting criteria can be changed using arrows and tabs

topas - process metrics

- USER
  - login name of the process owner
- PID
  - process ID of the process
- PPID
  - process ID of the parent process
- PRI
  - priority of the process or kernel thread
    - ▶ higher numbers mean lower priority.
- NI
  - nice value
    - ▶ used in calculating scheduling priority
- TIME
  - total execution time for the process
- CPU%
  - percentage of CPU usage

- Data RSS
  - real memory data (resident set) size of the process
    - ▶ in 4KB pages
- TEXT RSS
  - real memory text (resident set) size of the process
    - ▶ in 4KB pages
- PAGE SPACE
  - virtual working set size used by process
    - ▶ in 4KB pages
- PGFAULTS
  - number of I/O and other page faults
- COMMAND
  - command name truncated to 9 characters

topas - main screen with WLM active

```
Topas Monitor for host:     niller          EVENTS/QUEUES       FILE/TTY
Mon Mar 13 15:56:32 2000   Interval:  2     Cswitch      113    Readch   1853576
                                            Syscall     2510    Writech    49883
CPU    User%    Kern%    Wait%    Idle%     Reads        466    Rawin          0
cpu0    7.0      4.0      0.0     89.0       Writes        12    Ttyout       706
cpu1    1.0      8.0      0.0     91.0       Forks          0    Igets          0
cpu2    0.0      0.0      0.0    100.0       Execs          0    Namei          0
                                            Runqueue     0.0    Dirblk         0
                                            Waitqueue    0.0
Interf    KBPS    I-Pack   O-Pack    KB-In    KB-Out
lo0      100.4    45.7     45.7     50.2     50.2    PAGING            MEMORY
tr0        2.0     4.4      3.4      1.4      0.6    Faults       1    Real,MB      255
                                                    Steals       0    % Comp      81.0
Disk    Busy%     KBPS     TPS KB-Read KB-Writ       PgspIn       0    % Noncomp   19.0
hdisk0   0.0      0.0      0.0     0.0      0.0       PgspOut      0    % Client     3.0
hdisk1   0.0      0.0      0.0     0.0      0.0       PageIn       0
                                                     PageOut      0    PAGING SPACE
WLM-Class (Active)   CPU%     Mem%     Disk%         Sios         0    Size,MB      512
System                8       41       12                              % Used      33.6
Shared                1       24        9                              % Free      66.4
                                                    NFS      calls/sec
Name       PID     CPU%    PgSP    Class           ServerV2     0    Press:
topas     (35242)  3.0     0.3     System          ClientV2     0    "h" for help.
X         (3622)   1.4     44.4    System          ServerV3     0    "q" to quit.
notes     (25306)  1.3     123.3   System          ClientV3     0
cron      (5244)   0.0     0.5     System
ksh       (10674)  0.0     0.5     system
snmpd     (6966)   0.0     0.9     System
```

topas - full screen WLM process view

```
Topas Monitor for host:     mothra        Interval:    2      Wed Nov  8 12:30:54 2000
WLM-Class (Active)                      CPU%         Mem%      Disk-I/O%
System                                   0            0            0
Shared                                   0            0            0
Default                                  0            0            0
Unmanaged                                0            0            0
Unclassified                             0            0            0


============================================================================

                             DATA    TEXT   PAGE                    PGFAULTS
USER          PID  PPID PRI NI  RES    RES  SPACE      TIME CPU%  I/O  OTH COMMAND
root            1     0 108 20  197      9    180      0:24  0.0    0    0 init
root         1032     0  16 41    3   3374      3      0:00  0.0    0    0 lrud
root         1290     0  60 41    4   3374      4      0:02  0.0    0    0 xmgc
root         1548     0  36 41    4   3374      4      0:26  0.0    0    0 netm
root         1806     0  37 41   16   3374     16     13:25  0.0    0    0 gil
root         2064     0  16 41    4   3374      4      0:04  0.0    0    0 wlmsched
root         2698     1 108 20   14      2     14      0:00  0.0    0    0 shlap
root         3144     1 108 20   40      1     36      5:19  0.0    0    0 syncd
root         3362     0 108 20    4   3374      4      0:00  0.0    0    0 lvmbb
root         3666     1 108 20  135     23    123      0:00  0.0    0    0 errdemon
root         3982     0 108 20    4   3374      4      0:01  0.0    0    0 rtcmd
```

- Output from topas -W or "W" command
- Class focus can be changed by moving cursor over class and using "f" key
- Sorting criteria can be changed using arrows and tabs in both sections

topas - full screen disk view

```
Topas Monitor for host:    lambic       Interval:   2    Wed Jan 22 12:10:44 2003
================================================================================
Disk      Busy%   KBPS     TPS    KB-R    ART    MRT    KB-W    AWT    MWT    AQW    AQD
hdisk0     2.0    12.0     3.0     0.0    0.0    9.1    12.0    6.0   16.5    0.0    0.0
hdisk6     5.0    24.5     6.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk4    12.0    72.2    20.5    36.1    0.0    0.0     0.0    0.0   11.1    0.0    0.0
hdisk1     0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk7     4.5     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk3     0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0  147.8    0.0    0.0
hdisk5     0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk11    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk13    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk2     7.0    20.3     9.0     0.0    0.0    0.0     0.0    0.0    9.2    0.0    0.0
hdisk17    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
cd0        0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk12    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk15    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk14    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk18    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk21    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk20    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk19    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk22    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
hdisk16    0.0     0.0     0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0    0.0
```

main screen metrics

- Output from topas -D or "D" command
- Sorting criteria can be changed using arrows and tabs

topas - disk metrics

- Disk
  - name of the disk
- Busy%
  - percentage of time the disk was active
- KBPS
  - kilobytes read and written per second
  - sum of KB-Read and KB-Read
- TPS
  - number of transfers per second to the disk.
    - a transfer is an I/O request to the disk
    - multiple logical requests can be combined into a single I/O request
    - transfer request sizes are variable
- KB-Read
  - kilobytes read per second from the disk
- K -Writ
  - kilobytes written per second to the disk

`main screen metrics`

- AQD
  - average number of requests waiting to be sent to disk

- AQW
  - average queue wait time per request
- ART
  - average service time for read requests
- MRT
  - maximum service time for read requests
- AWT
  - average service time for write requests
- MWT
  - maximum serice time for write requests

- The default service and wait time unit is milliseconds
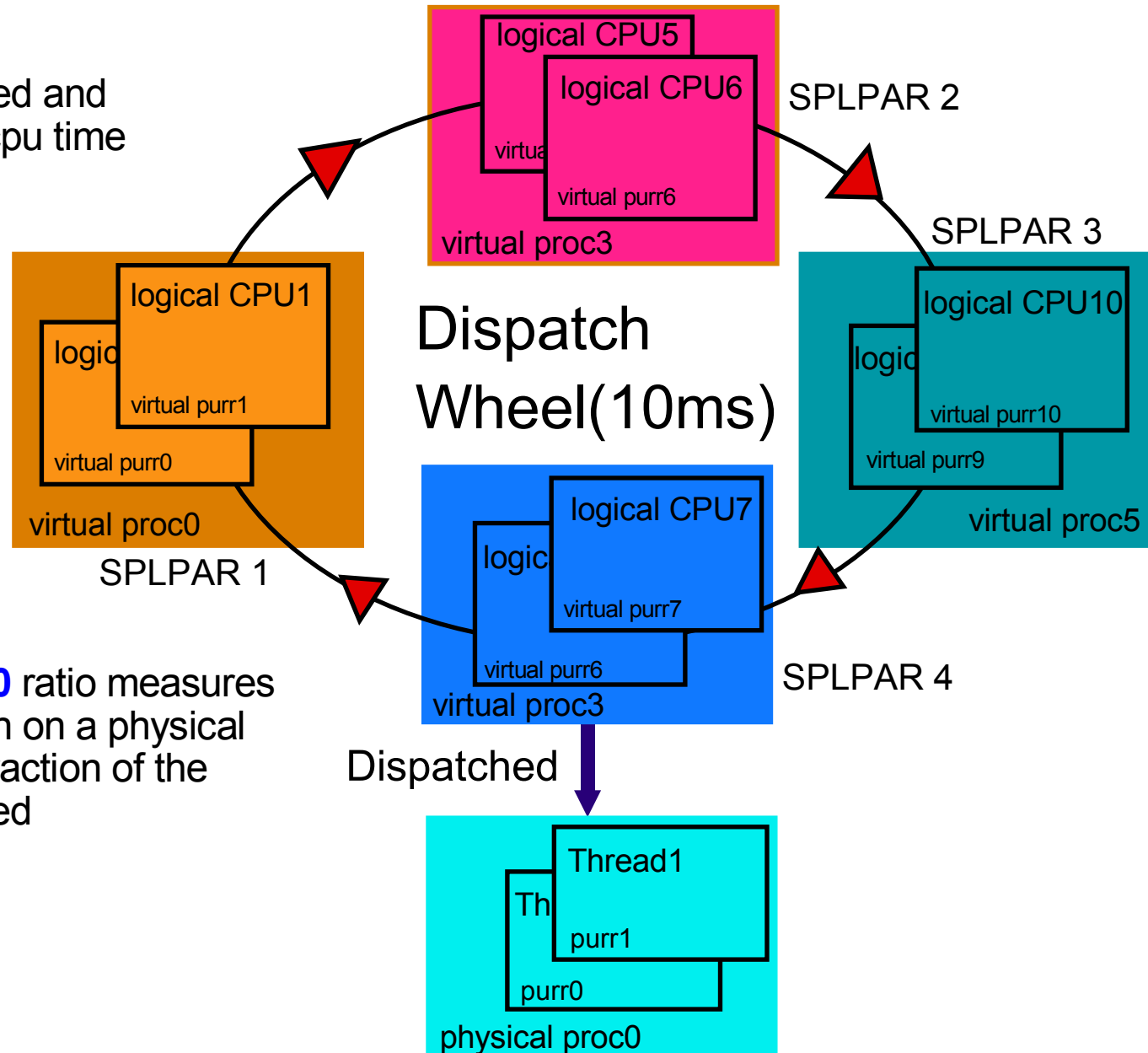 - the suffix indicates the unit of time.

PURR - Processor Utilization Resource Register

- New register implemented by Power5

- Each thread has its own PURR

- Units are the same as the Timebase and Decrementer registers

- For a given time interval, the sum of the PURR delta values for both threads is equal to deltaTimebase values

- PURR increments actually measure instruction dispatch cycles
  - at each cycle only one of the two purr registers gets incremented
    - ► either the thread dispatching instructions
    - ► or the thread who last dispatched instructions
  - provides a reasonable indicator of the split of the work between the two threads

- Also used by PHYP to virtualize time between micropartitions

Logical

CPU0

Logical

CPU1

Thread0
purr0

Thread1
purr1
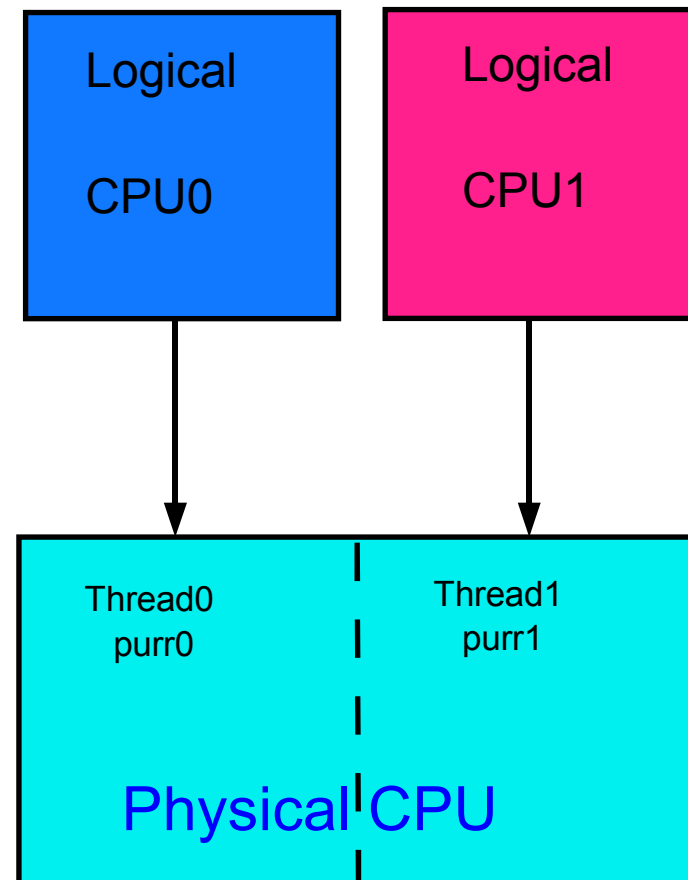
Physical CPU

Shared Processor LPAR with SMT

- Two time metrics
  - thread's PURR are virtualized and measure thread's physical cpu time
  - TimeBase still measures elapsed (wall clock) time

- Two PURR usages
  - **vpurr0/vpurr1** ratio measures relative SMT split between threads
  - **(vpurr0+vpurr1)/timebase0** ratio measures fraction of time partition1 ran on a physical processor, i.e. the relative fraction of the physical processor consumed

logical CPU5

logical CPU6

SPLPAR 2

virtua

virtual purr6

virtual proc3

SPLPAR 3

logical CPU1

logic

virtual purr1

Dispatch

virtual purr0

Wheel(10ms)

logical CPU10

logic

virtual purr10

virtual proc0

virtual purr9

SPLPAR 1

virtual proc5

logical CPU7

logic

virtual purr7

virtual purr6

SPLPAR 4

virtual proc3

Dispatched

Thread1

Th

purr1

purr0

physical proc0

SMT and SPLPAR - CPU consumption metrics

- CPU consumption monitoring
  - partition level
    - measured by monitoring virtualized PURR and absolute TimeBase register
      - delta PURR/delta TimeBase = physical processor consumed
  - shared pool level
    - measured by monitoring the "Pool Idle Count"
    - maintained by the hypervisor and represents clock ticks where a physical processor was idle
      - all partition entitlements satisfied
      - no partition to dispatch

- Metrics displayed
  - traditional cpu utilization breakdown (%user, %sys, %idle, %wait)
    - measured using PURR and reported relative to entitlement except when above cap
  - other shared mode only utilization metrics
    - Physical Processor Consumed (**physc**)
      - delta PURR/delta TimeBase
    - Percentage of Entitlement Consumed (**%entc**)
      - (physc/entitlement)*100

SMT and SPLPAR - additional metrics

- Hypervisor calls statistics
  - number of hcalls (**hcalls**)
    - ►total number of hcalls per second
      - currently includes voluntary context switch calls
  - percentage spent in hypervisor (**%hypv**)
    - ►total time spent in the hypervisor processing hcalls
    - ►reported as a percentage of cpu consumption
      - not comparable with %user, %sys, %idle and %wait except for uncapped partitions above entitlement
    - ►this is NOT a measure of hypervisor/partitioning overhead

- Additional shared mode only metrics
  - Available Pool Processors (**app**)
    - ►delta Pool Idle Count value converted into decimal number of processor
    - ►if the partition has the authority to retrieve information
  - number of virtual context switches per second (**vcsw**)
    - ►monitoring number of hardware context switches per second
  - number of phantom interrupts per second (**phint**)
    - ►monitoring number of interrupts per second

topas - main screen support for micro-partitions

- New cpu section metrics on physical processing resources consumed
  - Physc: amount consumed in fractional number of processors
  - %Entc: amount consumed in percentage of entitlement
  - automatically displayed when running in shared mode

```
Topas Monitor for host:     specweb8          EVENTS/QUEUES     FILE/TTY
Sat Mar 13 09:47:18 2004    Interval:  2       Cswitch     50    Readch        0
                                               Syscall     47    Writech      34
Kernel    0.0    |                          |  Reads        0    Rawin         0
User      0.0    |                          |  Writes       0    Ttyout       34
Wait      0.0    |                          |  Forks        0    Igets         0
Idle    100.0    |###########################|  Execs        0    Namei         1
Physc =  0.01                      %Entc=   1.2  Runqueue    0.0   Dirblk        0
                                                 Waitqueue   0.0

Network   KBPS    I-Pack   O-Pack    KB-In   KB-Out
en0       0.1      1.0      1.0       0.0      0.1   PAGING            MEMORY
lo0       0.0      0.0      0.0       0.0      0.0   Faults       0    Real,MB    8191
                                                    Steals       0    % Comp      5.4
Disk      Busy%      KBPS      TPS KB-Read KB-Writ   PgspIn       0    % Noncomp   1.6
hdisk0    0.0       0.0       0.0    0.0     0.0     PgspOut      0    % Client    1.6
hdisk2    0.0       0.0       0.0    0.0     0.0     PageIn       0
hdisk3    0.0       0.0       0.0    0.0     0.0     PageOut      0    PAGING SPACE
hdisk1    0.0       0.0       0.0    0.0     0.0     Sios         0    Size,MB     512
                                                                      % Used      0.6
Name           PID   CPU%  PgSp Owner          NFS (calls/sec) % Free      99.3
IBM.CSMAg     13180   0.0   1.6 root           ServerV2     0
syncd          9366   0.0   0.5 root           ClientV2     0    Press:
prngd         22452   0.0   0.3 root           ServerV3     0    "h" for help
psgc           2322   0.0   0.0 root           ClientV3     0    "q" to quit
pilegc         2580   0.0   0.0 root
```

- CPU utilization metrics are automatically calculated using new PURR-based data and formula when running in SMT or shared mode

topas - LPAR screen

- Split screen accessible from -L or the "L" command
  - upper section shows a subset of **lparstat** metrics
  - lower section shows sorted list of logical processor with **mpstat** columns

```
Interval:    2          Logical Partition: aix      Sat Mar 13 09:44:48 2004
Poolsize: 3.0                    Shared SMT ON          Online Memory:    8192.0
Entitlement: 2.5                   Mode: Capped         Online Logical CPUs:   4
                                                        Online Virtual CPUs:   2
%user  %sys  %wait  %idle physc  %entc %lbusy    app   vcsw  phint %hypv hcalls
 47.5  32.5    7.0   13.0   2.0   80.0  100.0    1.0    240    150   5.0   1500
==============================================================================
logcpu minpf majpf   intr   csw   icsw runq lpa   scalls usr sys  wt idl  pc lcsw
cpu0    1135   145    134    78     60    2  95    12345   10  65  15  10 0.6  120
cpu1     998   120    104    92     45    1  89     4561    8  67  25   0 0.4  120
cpu2    2246   219    167   128     72    3  92    76300   20  50  20  10 0.5  120
cpu3    2167   198    127    62     43    2  94     1238   18  45  15  22 0.5  120
```

- Notable metrics
  - %hypv and hcalls: percentage of time in hypervisor and number of calls made
  - pc: fraction of physical processor consumed by a logical processor
  - additional shared mode only metrics
    - ►Poolsize: number of processors in LPAR pool this partition belongs to
    - ►app: available pool processors
    - ►physc: number of physical processor(s) consumed
    - ►%entc: percentage of entitlement consumed
    - ►lcsw and vcsw: logical and virtual context switches
    - ►phint: number of phantom interrupts

topas - CEC monitoring screen

- Split screen accessible from -C or the "C" command
  - upper section shows CEC-level metrics
  - lower sections shows sorted lists of shared and dedicated partitions

```
Topas CEC Monitor                   Interval:   10              Thu Jul 28 17:04:57 2005
Partitions        Memory (GB)                   Processors
Shr:  3           Mon:24.6  InUse: 2.7          Shr: 1.5  PSz:  3   Shr_PhysB: 0.27
Ded:  3           Avl:   -                      Ded:  5  APP:  2.6  Ded_PhysB: 2.70

Host            OS   M Mem InU Lp   Us Sy Wa Id  PhysB   Ent   %EntC Vcsw  PhI
-----------------------------------------------shared----------------------------------
ptoolsl3        A53 c 4.1 0.4   2  14  1   0 84   0.08   0.50   15.0  208    0
ptoolsl2        A53 C 4.1 0.4   4  20 13   5 62   0.17   0.50   36.5  219    5
ptoolsl5        A53 U 4.1 0.4   4   0  0   0 99   0.02   0.50    0.1  205    2

----------------------------------------------dedicated--------------------------------
ptoolsl1        A53 S 4.1 0.5   4  20 10   0 70   0.60
ptoolsl4        A53   4.1 0.5   2 100  0   0  0   2.00
ptoolsl6        A52   4.1 0.5   1   5  5  12 88   0.10
```
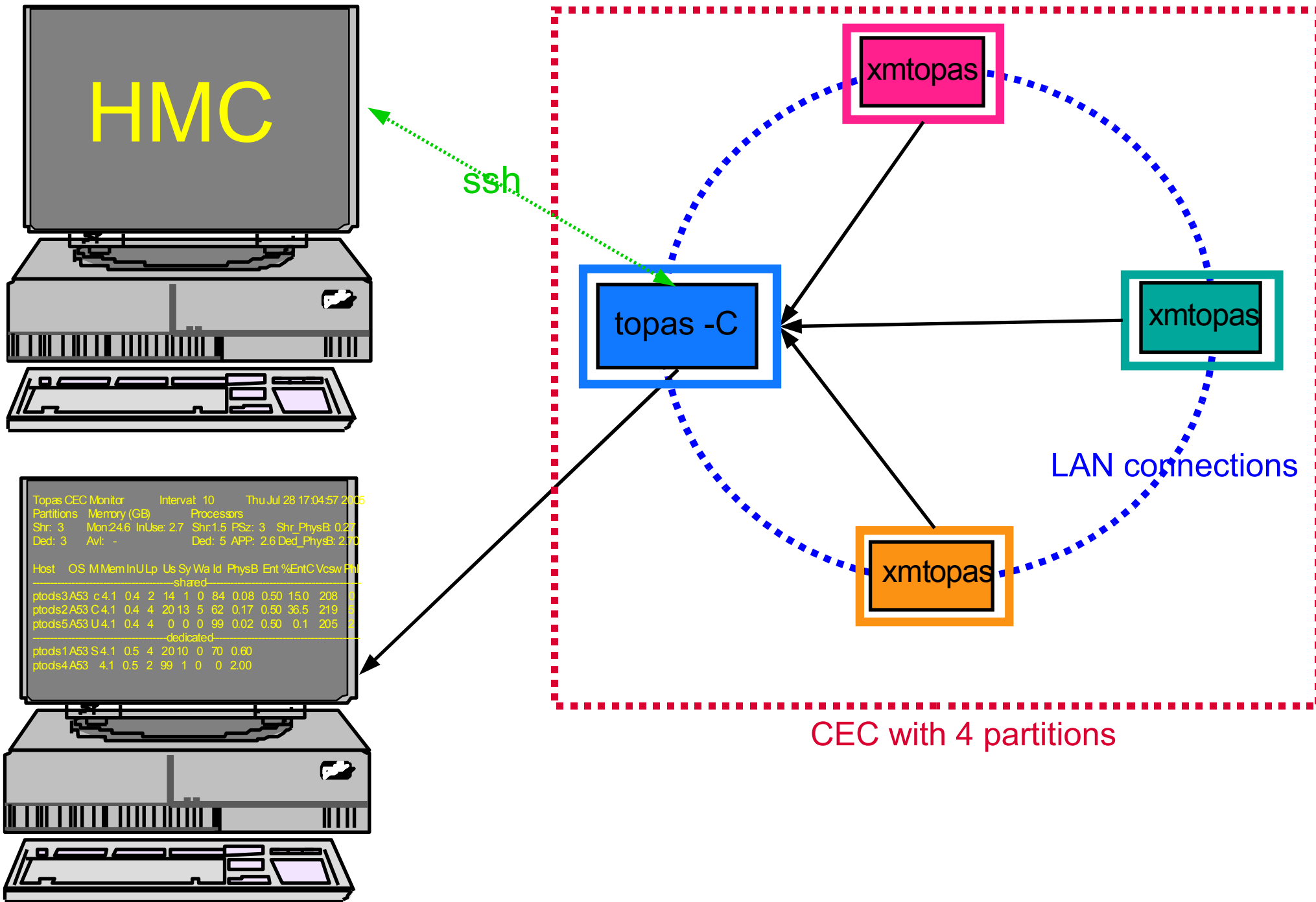
- CEC configuration info retrieved from HMC or specified on command line
- Mode column
  - c means capped, C – capped with SMT on
  - u means uncapped – U uncapped with SMT on
  - S means SMT on
- Uses the new **xmtopas** daemon started by inetd

topas - CEC monitoring data collection setup

HMC

ssh

xmtopas

topas -C

xmtopas

xmtopas

LAN connections

CEC with 4 partitions

```
Topas CEC Monitor        Interval: 10        Thu Jul 28 17:04:57 2005
Partitions   Memory (GB)              Processors
Shr:  3      Mon:24.6 InUse: 2.7  Shr:1.5 PSz: 3   Shr_PhysB: 0.27
Ded:  3      Avl:  -                Ded: 5 APP:  2.6 Ded_PhysB: 2.70

Host    OS M Mem InU Lp  Us Sy Wa Id  PhysB Ent %EntC Vcsw PhI
-----------------------------shared-------------------------------
ptods3 A53 c 4.1  0.4  2  14  1  0 84  0.08 0.50 15.0  208   0
ptods2 A53 C 4.1  0.4  4  20 13  5 62  0.17 0.50 36.5  219   5
ptods5 A53 U 4.1  0.4  4   0  0  0 99  0.02 0.50  0.1  205   2
----------------------------dedicated-----------------------------
ptods1 A53 S 4.1  0.5  4  20 10  0 70  0.60
ptods4 A53   4.1  0.5  2  99  1  0  0  2.00
```

topas - CEC configuration information

- Command line options (-o flag)
  - availmem
    - ►total memory size allocated in MB
  - unavailmem
    - ►total memory size unallocated in MB
  - availcpu
    - ►total cpus allocated on the HMC
  - unavailcpu
    - ►total cpus unallocated on the HMC
  - partitions
    - ►number of partitions defined on the HMC
  - reconfig
    - ►reconfiguration interval in sec 30/60(default)/90/120/180/240/300
  - poolsize
    - ►defined Pool Size required if HMC PUA restricts access

- If some partitions are on different subnetworks
  - create a file named **Rsi.hosts** with one fully qualified hostname or IP address per line
  - put file either in $HOME or in **/etc/perf**

topas - partitions and HMC connection setup

- SSH setup
  - install OpenSSH on the partition
  - enable remote command support on the HMC for user hscroot to allow ssh connections to be opened from the partition
  - configure ssh on the HMC to not require a password for the HMC user hscroot when queried from the selected partition
    - ► requires the .ssh/authorized_keys2 on the HMC for user login hscroot
  - run
    - ► ssh -l hscroot hmc_address date
  - from the partition to confirm whether the date is displayed without requiring that a password be entered

- Topas setup (-o flag)
  - managedsys
    - ► managed system name under which this partition is configured
  - hmc
    - ► HMC name under which this partition is configured, available via:
      - lsrsrc -dx IBM.ManagementServer Hostname

topas - CEC and individual partition metrics

- **Partition totals**
  - Shr
    - Number of shared partitions
  - Ded
    - Number of dedicated partitions
- **Memory (in GB)**
  - Mon
    - Monitored partitions total memory
  - Avl
    - Memory available to partition set
  - InUse
    - Memory in use on monitored partitions
- **Processor**
  - Shr
    - Number of shared processors
  - Ded
    - Number of dedicated processors
  - PSz
    - Active physical CPUs in the shared processor pool being used by this LPAR
  - APP
    - Available physical processors in the shared pool
  - Shr_PhysB
    - Shared Physical Busy
  - Ded_PhysB
    - Dedicated Physical Busy

- **Partition metrics**
  - Host
    - Hostname
  - OS
    - Operating system level
  - M
    - Mode
      - C - SMT enabled & capped
      - c - SMT disabled & capped
      - U - SMT enabled & uncapped
      - u - SMT disabled & uncapped
      - S - SMT enabled (dedicated partitions)
  - Mem, InU
    - Total and in use memory in GB.
  - Lp
    - Number of logical processors
  - Us, Sy, Wa, Id
    - Percentage of CPU in user mode, kernel mode, spent waiting for IO, and idle.
  - PhysB
    - Physical Processors Busy
  - Ent, %Entc
    - Entitlement and Percent Entitlement consumed
  - Vcsw
    - Virtual context switches per second
  - PhI
    - Phantom interrupts per second

shared mode only

topas - CEC panel commands

- g
  - toggles the top global section of the panel between
    - brief listing
    - detailed listing
    - off
- d
  - toggles the dedicated partition section on and off.
- s
  - toggles the shared partition section on and off.
- r
  - forces topas to search for HMC configuration changes if a connection is available. This includes the discovery of new partitions, processors, or memory allocations.

Automatic Performance Metric recording

- Introduced in 5.3 TL 4
  - uses xmwlm daemon
  - automaticaly started from inittab
  - initially kept 2 days worth of data, but changing to 7 days in 5.3 TL5
  - recordings include most of topas data
    - ►except process and WLM data

- New (5.3 TL5) topas -R option records topas -C metrics (CEC-wide data)
  - works independently and in parallel from topas real-time monitors
  - must be turned on manually in one of the partitions in CEC
    - ►via configuration script which adds line in inittab

      ```
      # /usr/lpp/perfagent/config_topas.sh add
      ```

- topasout
  - postprocessing tool for recordings
  - WLE reports
  - text reports (5.3 TL5)
    - ►include both local data and CEC-wide data
    - ►options include detailed and summary
  - spreadsheet and csv formats
  - nmon_analyzer format (5.3 TL5)

Automatic Performance Metric recording*(cont)*
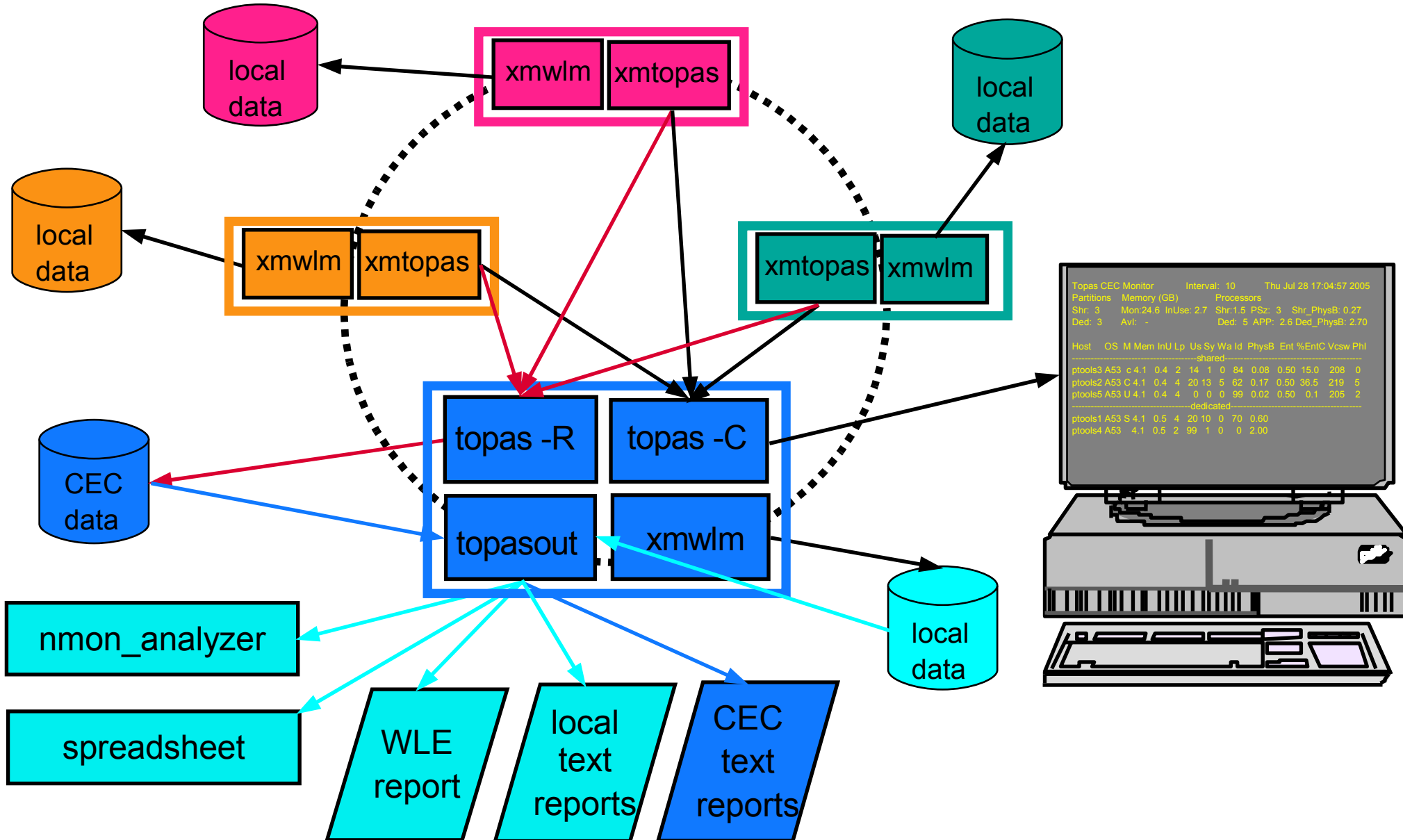
topasout - CEC summary report

- Example with configuration change

```
Report: Topas CEC Summary --- hostname: ptoolsl1                    version:1.0
Start:02/09/06 00.00.00 Stop:02/09/06 23.55.00  Int:  5 Min  Range:1440 Min
Partition Mon:  7  UnM:  1  Shr:  4   Ded:  3   Cap:  3   UnC:  1
            -CEC------    -Processors-------------------- -Memory (GB)------------
Time        ShrB   DedB   Mon UnM Avl UnA Shr Ded PSz APP   Mon   UnM   Avl   UnA   InU
00.05.00    3.2    1.1     5   2   7   1   4   3   2   1   16.0  0.0   32.0  0.0   8.1
00.10.00    2.9    0.9     5   2   7   1   4   3   2   1   16.0  0.0   32.0  0.0   8.3
00.15.00    2.1    1.3     5   2   7   1   4   3   2   1   16.0  0.0   32.0  0.0   8.5

...


configuration change at 02.15.00
Partition Mon:  8  UnM:  0  Shr:  4   Ded:  4   Cap:  3   UnC:  1
            -CEC------    -Processors-------------------- -Memory (GB)------------
Time        ShrB   DedB   Mon UnM Avl UnA Shr Ded PSz APP   Mon   UnM   Avl   UnA   InU
02.15.00    3.1    2.5     7   0   7   1   4   5   2   1   18.0  0.0   32.0  0.0   9.1
02.20.00    1.9    1.5     7   0   7   1   4   5   2   1   18.0  0.0   32.0  0.0   6.8
02.25.00    2.0    3.3     7   0   7   1   4   5   2   1   18.0  0.0   32.0  0.0   7.8

...
```

topasout - detailed CEC report

```
Report: Topas CEC Detailed --- hostname: ptoolsl1                    version:1.0
Start:05/02/06 07.00.00  Stop:05/02/06 17.00.00  Int:05 Min  Range:600 Min
Time: 07.00.00 -----------------------------------------------------------------
Partition Info      Memory (GB)        Processors
Monitored  :  8     Monitored  : 0.0   Monitored  :  7   Shr Physical Busy:  2.2
UnMonitored:  -     UnMonitored: 0.0   UnMonitored:  0   Ded Physical Busy:  0.4
Shared     :  3     Available  :32.0   Available  :  7
Dedicated  :  2     UnAllocated:  -    UnAllocated:  1   Hypervisor
Capped     :  1     Consumed   : 8.7   Shared     :  4   Virt. Context Switch:332
Uncapped   :  2                        Dedicated  :  3   Phantom Interrupts  :  2
                                       Pool Size  :  2
                                       Avail Pool :  1

Host          OS   M Mem InU Lp  Us Sy Wa Id  PhysB   Ent   %EntC Vcsw PhI
-------------------------------------shared-------------------------------------
ptools1       A53 u 1.1 0.4  4  15  3  0 82   1.30  0.50  22.0  200   5
ptools5       A53 U  12  10  1  12  3  0 85   0.20  0.25   0.3  121   3
ptools3       A53 C 5.0 2.6  1  10  1  0 89   0.15  0.25   0.3  52    2
-------------------------------------dedicated----------------------------------
ptools4       A53 S 0.6 0.3  2  12  3  0 85   0.60
ptools6       A52   1.1 0.1  1  11  7  0 82   0.50
ptools8       A52   1.1 0.1  1  11  7  0 82   0.50
Time: 07.05.00 -----------------------------------------------------------------
Partition Info      Memory (GB)        Processors
Monitored  :  8     Monitored  : 0.0   Monitored  :  7   Shr Physical Busy:  2.2
UnMonitored:  -     UnMonitored: 0.0   UnMonitored:  0   Ded Physical Busy:  0.4
Shared     :  3     Available  :32.0   Available  :  7
Dedicated  :  2     UnAllocated:  -    UnAllocated:  1   Hypervisor
Capped     :  2     Consumed   : 8.7   Shared     :  4   Virt. Context Switch:332
Uncapped   :  2                        Dedicated  :  3   Phantom Interrupts  :  2
                                       Pool Size  :  2
                                       Avail Pool :  1

Host          OS   M Mem InU Lp  Us Sy Wa Id  PhysB   Ent   %EntC Vcsw PhI
-------------------------------------shared-------------------------------------
ptools1       A53 u 1.1 0.4  4  15  3  0 82   1.30  0.50  22.0  200   5
ptools5       A53 U  12  10  1  12  3  0 85   0.20  0.25   0.3  121   3
ptools3       A53 C 5.0 2.6  1  10  1  0 89   0.15  0.25   0.3  52    2
-------------------------------------dedicated----------------------------------
ptools4       A53 S 0.6 0.3  2  12  3  0 85   0.60
ptools6       A52   1.1 0.1  1  11  7  0 82   0.50
ptools8       A52   1.1 0.1  1  11  7  0 82   0.50
Time: 07.10.00 -----------------------------------------------------------------
```

topasout - summary local report

- Dedicated partitions

```
Report: System Summary - hostname: ptoolsl1                    version 1.0


Start:12/20/05 14.00.00 Stop:12/20/05 15.00.00  Int: 5 Min  Range: 60 Min
Mem: 16.2 GB  Dedicated SMT:OFF  Logical CPUs:  2
Time       InU  Us Sy Wa Id  PhysB   RunQ WtQ CSwitch Syscall PgFault
----------------------------------------------------------------------
14.00.00  21.1 11  8  0 81    0.2     1   0    3432    5050      17
14.05.00  21.1 16  5  0 79    0.3     1   0     532    3104      14
14.10.00  21.2 13  7  0 20    0.2     1   0     652    4326      13
```

- Shared partitions

```
Report: System Summary - hostname: ptoolsl1                    version 1.0


Start:12/21/05 10.00.00 Stop:12/21/05 11.00.00  Int: 5 Min  Range: 60 Min
Psize:1.0  Mem: 16.2 GB  Shared SMT:OFF  Logical CPUs:  2
Time       InU  Us Sy Wa Id PhysB   Ent   %EntC RunQ WtQ CSwitch Syscall PgFault
--------------------------------------------------------------------------------
10.00.00  21.1 11  8  0 81   0.2    0.5   23.2   1   0    3432    5050      17
10.05.00  21.1 16  5  0 79   0.3    0.5   25.0   1   0     532    3104      14
10.10.00  21.2 13  7  0 20   0.2    0.5   23.4   1   0     652    4326      13
```

topasout - detailed local report

```
Report: System Detailed --- hostname: ptoolsl1                   version 1.0
Start:12/21/05 10.00.00 Stop:12/21/05 11.00.00  Int: 5 Min  Range: 60 Min
Time: 10.00.00 --------------------------------------------------------------
CPU           UTIL         MEMORY        PAGING        EVENTS/QUEUES  NFS
Kern  12.0    PhyB   0.7   Sz,GB  16.0   Sz,GB   4.0   Cswth    3213  SrvV2   32
User   8.0    Ent    0.5   InU     4.3   InU     2.3   Syscl   43831  CltV2   12
Wait   0.0    EntC  15.2   %Comp   3.1   Flt     221   RunQ        1  SrvV3   44
Idle  78.0    LP       4   %NonC   9.0   Pg-I     87   WtQ         0  CltV3   18
SMT    ON     Mode   Shr   %Clnt   2.0   Pg-O     44   VCSW     1214


Network  KBPS      I-Pack   O-Pack      KB-I     KB-O
en0       0.6        7.5      0.5        0.3      0.3
en1      22.3      820.1    124.3      410.0     61.2
lo0       0.0        0.0      0.0        0.0      0.0


Disk    Busy%      KBPS      TPS      KB-R     KB-W
hdisk0    0.0       0.0      0.0      0.0      0.0
hdisk1    0.0       0.0      0.0      0.0      0.0
Time: 10.05.00 --------------------------------------------------------------
CPU           UTIL         MEMORY        PAGING        EVENTS/QUEUES  NFS
Kern  12.0    PhyB   0.7   Sz,GB  16.0   Sz,GB   4.0   Cswth    3213  SrvV2   32
User   8.0    Ent    0.5   InU     4.3   InU     2.3   Syscl   43831  CltV2   12
Wait   0.0    EntC  15.2   %Comp   3.1   Flt     221   RunQ        1  SrvV3   44
Idle  78.0    LP       4   %NonC   9.0   Pg-I     87   WtQ         0  CltV3   18
SMT    ON     Mode   Shr   %Clnt   2.0   Pg-O     44   VCSW     1214


Network  KBPS      I-Pack   O-Pack      KB-I     KB-O
en0       0.6        7.5      0.5        0.3      0.3
en1      22.3      820.1    124.3      410.0     61.2
lo0       0.0        0.0      0.0        0.0      0.0


Disk    Busy%      KBPS      TPS      KB-R     KB-W
hdisk0    0.0       0.0      0.0      0.0      0.0
hdisk1    0.0       0.0      0.0      0.0      0.0
```

topasout - I/O summary reports

- Disk report

```
Report: Total Disk I/O Summary - hostname: ptoolsl1                     version:1.0
Start:04/25/06 00.00.00  Stop:04/26/06 00.00.00  Int:05 Min  Range:1440 Min
Mem: 8.0 GB  Dedicated SMT:ON  Logical CPUs:16
Time          InU    PhysB    %Bsy  MBPS    TPS   MB-R   MB-W
--------------------------------------------------------------------------------
00.05.00      6.5    12.50    45.5 120.5 300.1 100.1   20.4
00.10.00      6.7    13.40    55.0 240.0 320.2 240.0    0.0
00.15.00      7.0    14.70    60.4 160.2 350.3  40.1 120.1
00.20.00      7.4    15.50    72.3 200.7 410.5  20.3 180.4
```

- LAN report

```
Report: Total LAN I/O Summary - hostname: ptoolsl1                      version:1.0
Start:03/12/06 17.15.00  Stop:03/12/06 20.30.00  Int:05 Min  Range: 195 Min
Psize:1.0  Mem: 16.2 GB   Shared SMT:OFF  Logical CPUs:  2
Time          InU    PhysB    MBPS   I-Pack O-Pack MB-I   MB-O Rcvdrp Xmtdrp
--------------------------------------------------------------------------------
17.15.00      3.2    6.30     20.0   310.5 120.2  16.2    3.8 120    160
17.20.00      3.3    6.45     22.3   220.3 225.7  11.1   11.2 118    165
17.25.00      3.2    6.15     18.5   275.6 158.0  11.6    6.9 121    162
17.30.00      3.4    6.55     19.4   270.2 156.9  11.3    6.1 124    154
```

topas - PTX affinity

- topas -C and -R
  - can monitor AIX 5.2 partitions if xmservd is running
- xmtopas is derived from xmservd
 - both used for real-time monitoring

- xmwlm is derived from xmtrend
  - collect "trend" data
    - ► average values
      - also min, max stdev
  - default mode collects average WLM performance metrics
    - ► viewed with wlmmon
      - free version of wlmperf which is part of PTX
  - -L option collects topas local data

- topas -R acts as DDS
  - Dynamic Data Supplier
  - "exports" aggregated CEC metrics to PTX name space
    - ► can be monitored with any of the PTX GUIs

PTX refresh - enhanced virtualization support

- Adds sample consoles for easy physical machine monitoring
  - skeleton console to display real-time aggregated data
    - only needs hostname of partition running topas -R to instantiate
  - skeleton console to monitor a set of partitions
    - only needs partitions hostnames to instantiate
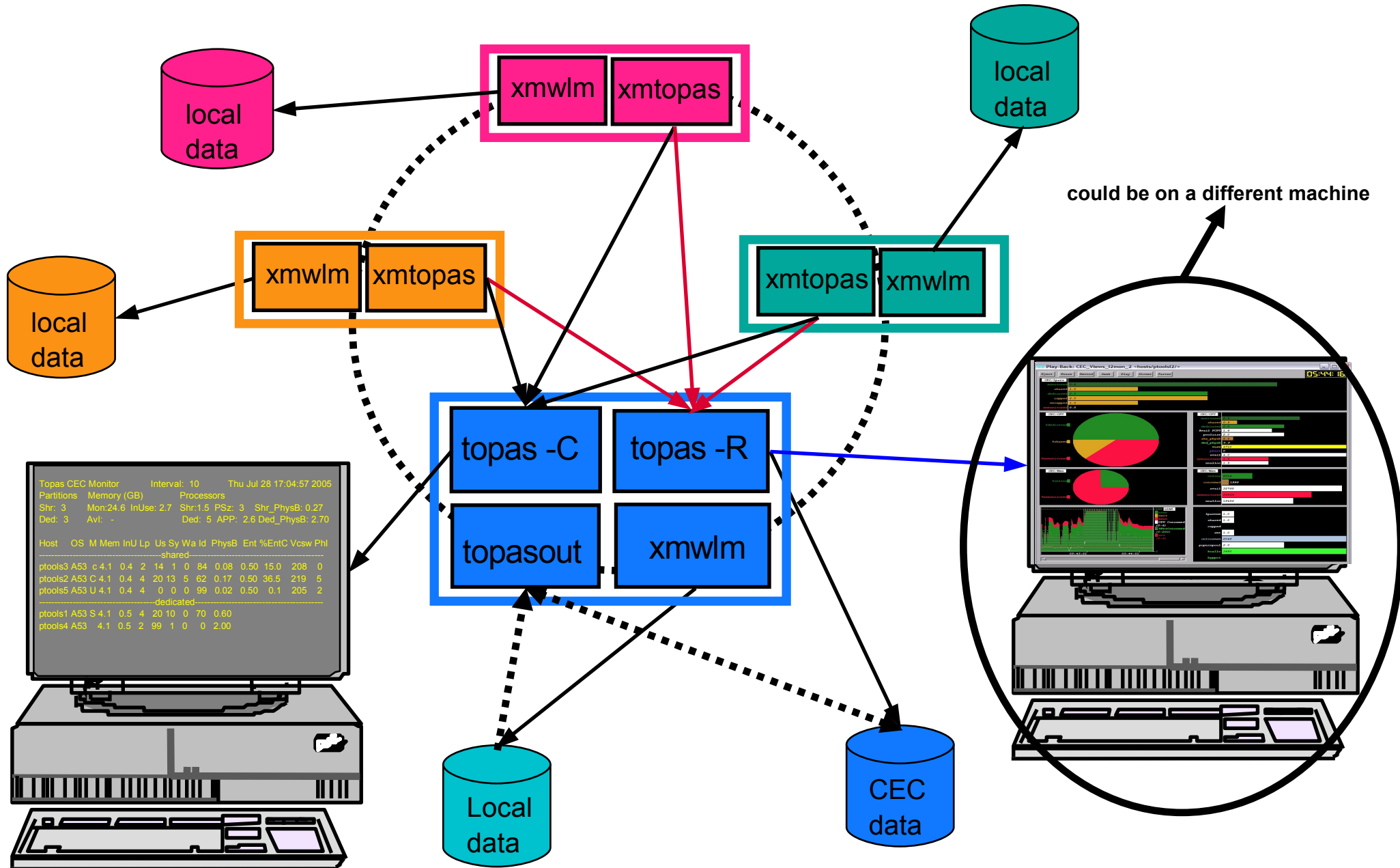   - includes 5.2 partitions running PTX agent
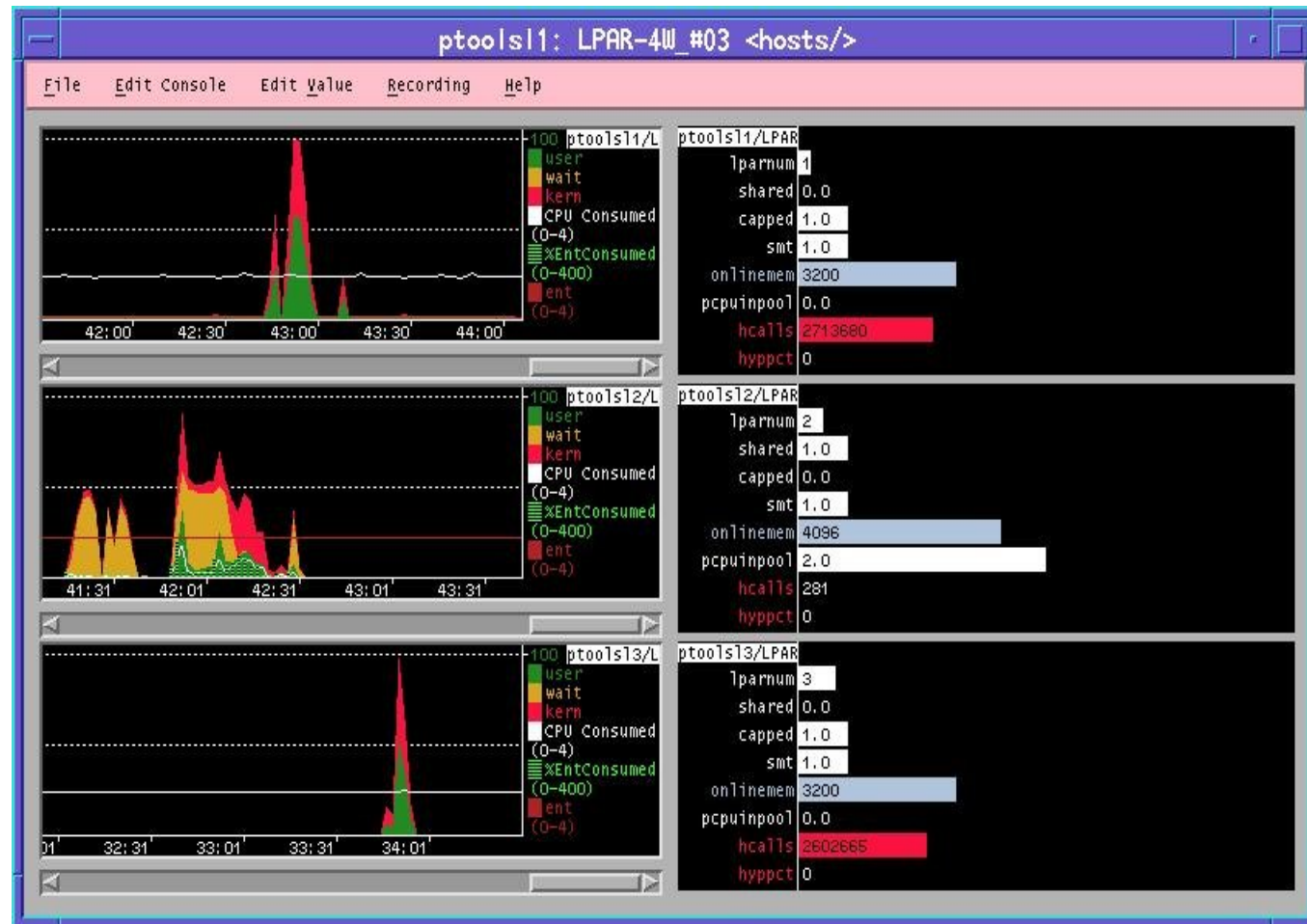
- Provides easily customizable solution
  - only list of partitions hostnames is needed to instantiate fully functional physical machine monitoring set of consoles
  - includes standard PTX attributes
    - recordings/playback
    - reports

PTX aggregated metrics viewing console
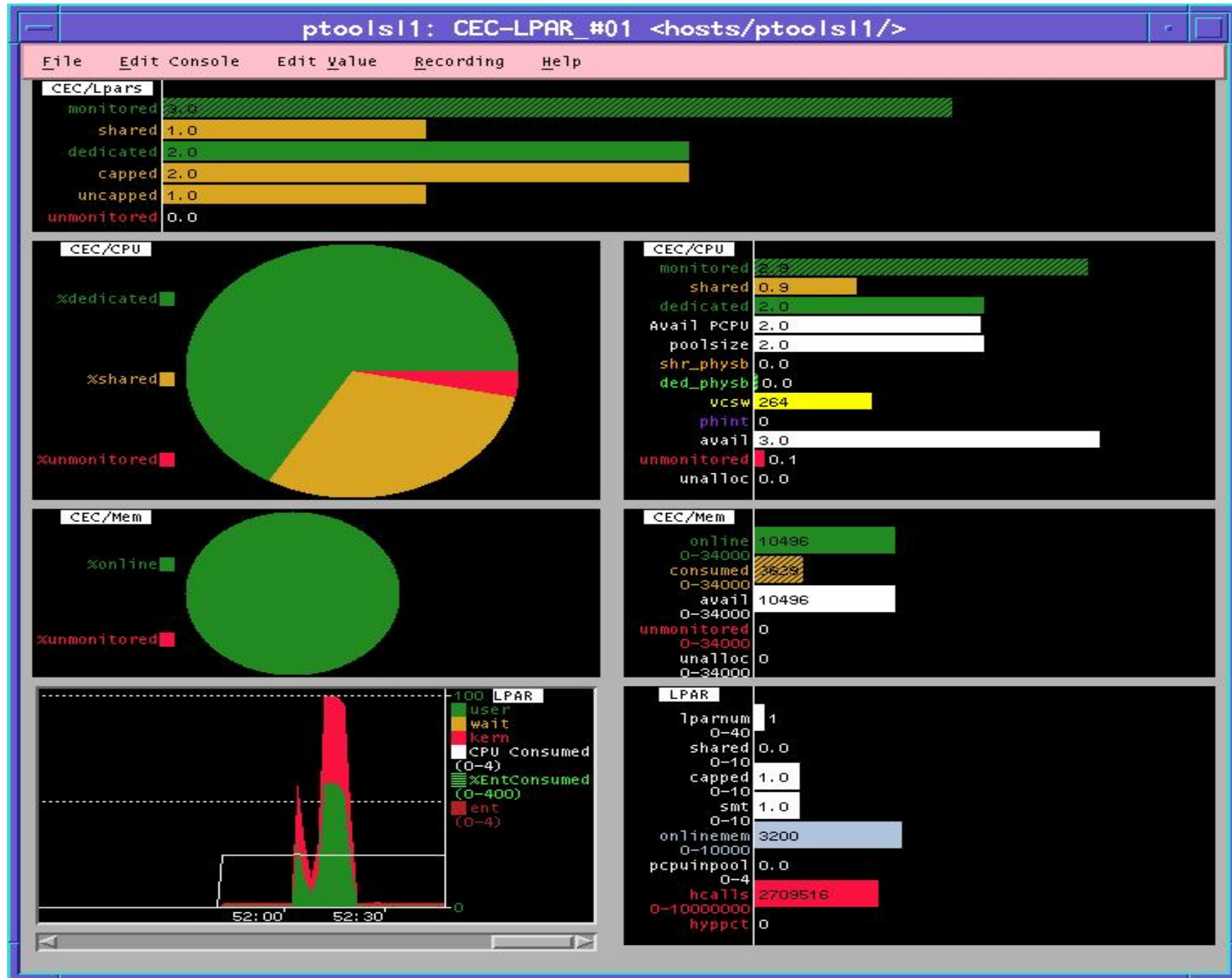
PTX aggregated metrics monitoring

Future plans

- VIOS support
  - topas -C suports VIOS partition with 08/06 code level
  - plan to add disk client/server view, recording and reports

- smit
  - configure Rsi.hosts
  - configure recordings (intervals, location)
  - generate reports

- topasout
  - support for WLE and WLM data text reports
  - rrdtools format

- nmon integration
  - merge real-time monitors
    - ► one binary with dual personalities
  - extend linkage with nmon_analyzer
  - fully support nmon_analyzer
    - ► similar to perfpmr support

Thank You!

charts to be available at:

http://www.ibm.com/developerworks/blogs/page/Systemptechuniv