開放的 열린 مفتوح libre मुक्त ಮ<u>ುಕ್ತ</u> livre libero ముక్త 开放的 acık open nyílt פתוח オープン livre ανοικτό offen otevřený öppen открытый வெளிப்படை



Introduction to Performance Monitoring Solaris POD – Performance, Observability & Debugging

Haim Tzadok, Cyril Plisko Grigale LTD

opensolaris

USE 🔆 IMPROVE (3) EVANGELIZE

Why Tuning ?



Alright, alright, you've won your bet: You can lift me with one hand...

opensolaris



Why Tuning ?

- Meet SLA(Service Level Agreement) standards
- Matching Workload to system capacity and configuration
- Maximizing consistency and throuput
- Reducing user response time
- Redeploying or balancing the load on system resource

opensolaris

Performance Terms and Definitions

- (R) Response Time
 - The time to complete a required operation
- (S) Service Time

- The time ot tales to serve an operation

• (W) Wait Time

- The time a process spent waiting for system resources

- Queueing (R=S+W)
 - Prediction of response time based on queueing model
- Bandwidth
 - Maximum signal(data) that can pass in a fixed period of time

USE improve (C) Evangelize



Terms and Definitions

- Throughput
 - The average amount of coherent data in a specific time interval
- Utilization
 - The precentage of elapsed time a component spent processing or serving a request

USE improve (3) Evangelize

General Methods and Approaches





CPU

- Finger rules:
- Vmstat kthr
 - run-queue(r) > 0 means process contention
 - block(b) >0 means processes are waiting for system resources.
 - w >0 means processes are swapped out (very bad)
- Vmstat cpu colomn
 - Good condition: cpu usr>2*sys
 - Bad condition: cpu usr<=2*sys
 - Check for system contention that can be derived from any other system resource.

CPU-cont

- Prstat load average
 - Load average should be not higher than 4*n-cores

USE IMPROVE (C) EVANGELIZE

- Prstat -L shows per thread statistics
- Mpstat Xcalls
 - Xcalls signaling between cpu's
 - Higher xcalls mean process thrashing between cpu's
 - Use: dtrace -n ':::xcalls {trace(execname)}' to find thrashing process.



Scheduling

- Scheduling classes in Solaris
 - TS Time sharing class(the original unix scheduling concept)
 - Sys (System class)
 - FX Fix priority class
 - IA Interactive class for use with interactive process
 - RT soft realtime class for high predictive response processes
 - FSS Fair Share Scheduler for using in zones and restraint resource management.
- dispadmin -l
 - Show which scheduling classes are activated on your system

Scheduling - cont

- priocntl a command to change process scheduling class and priority
- dispadmin -c TS -g show TS scheduling table and time quantum

USE improve (C) Evangelize

USE * IMPROVE (3) EVANGELIZE

Virtual Memory

- Anonymous pages pages that hold metadata not backed up by file-system.
- Page sizes: (use: pagesize -a)
 - Intel cpu contains 2 page sizes
 - Sparc cpu contains 4 page sizes
- Vmstat memory
 - Free free physical mem > $1/16^{th}$ of physical mem.
- Swap devices for large memory systems: consider dividing swap devices to several devices (kernel will perform striping on all swap devices), which result in faster anonpaging activity

Processes

- Every process has a dedicated virtual memory area containing: text, data, heap and stack
- Text contains executable code
- Data contains data produced at compilation time.
- Heap contains data produced at run-time.
- Stack contains process and threads function call and event activities



USE IMPROVE (C) EVANGELIZE

Process memory area illustration

Processes - cont

- P-commands /usr/proc/bin a lot of process analysis commands.
- For processes using memory intensive use: you can try tune page size using the command: ppgsz -o heap=256m

USE improve (C) Evangelize

I/O

- MPXIO I/O multi pathing can be performed on SAS and Fibre-channel devices for greater redundancy, load-balancing and throuput.
- iostat a command that shows service time and response time.
 - iostat -xznM (x -extended statistics, z filter zero values lines, n – print cxtxdx device names, M – show size in megabytes
- lostat -
 - $asvc_t disk/lun service time should be <=50ms$
 - wsvc_t disk/lun wait time(on non-empty queue).
 should be = 0ms

USE improve (C) Evangelize

Network

- Network multi-pathing
 - IPMP (Solaris 8 and above)
 - Link Aggregation(Solaris 10 and above)
- netstat -i
 - Colls number of collisions. Should be =0
 - Input Errs means some one is producing garbage on our network.
 - Output Errs means we are producing garbage on the network.
- netstat -s -p tcp
 - Retransmissions tcpRetransBytes < 5%



Caches

- Hardware and Software Caches
- HW caches -
 - L2 E-cache viewed by: cpustat
 - TLB(translation lookaside buffer) caches pages translations inside the CPU – viewed by: trapstat -t (only on SPARC systems)
- SW caches -
 - DNLC directory naming lookup cache viewed by: vmstat -s(total name lookups)
 - Buffercache file-system caching viewed by sar -b
 - Filesystem cache dynamic auto-resizing



Recommended Reference materials

- Solaris Internals books www.solarisinternals.com
- RICHPse se toolkit
- http://www.sunfreeware.com/setoolkit.html
- Dtrace toolkit

http://opensolaris.org/os/community/dtrace/dtra





Thank you !