PTR/INT Oracle Solaris 11 Internals

Infrastructure

DURATION	MODULES	COURSE CODE
5 Days	**11 Lectures**	**—**

Course Overview

This course will be available as Seminar-only, starting Jul 1, 2023.

What You Will Learn

Introduction

- Overview
- Course Goals
- Course Agenda: Day 1
- Course Agenda: Day 2
- Course Agenda: Day 3
- Course Agenda: Day 4
- Course Agenda: Day 5
- Introductions
- Your Learning Center

Oracle Solaris 11 Operating System: Introduction

- Objectives
- Lesson Agenda
- Operating System Basics
- Lesson Agenda
- Defining Processes
- Lesson Agenda
- SPARC 32-Bit Address Space
- SPARC 64-Bit Address Space
- x86 32-Bit Address Space: 64-Bit OS
- x86 64-Bit Address Space: 64-Bit OS
- Lesson Agenda
- Process Structures

- Lesson Agenda
- Kernel Mode Entry
- Lesson Agenda
- System Calls
- Interrupts
- Lesson Agenda
- clock() Routine
- LWP Accounting Scalability
- LWP Accounting
- Callout Queue
- callout Structure
- Callout Scalability
- Lesson Agenda
- Kernel Memory Allocator
- Lesson Agenda
- SunOS Evolution
- Solaris 9 +10: Features
- Solaris 11 OS: New Features
- Lesson Agenda
- Tools
- adb and kadb Tools
- mdb Tools
- kmdb Tools
- DTrace
- Lesson Agenda
- Probe Descriptions and Clauses
- Practice 2 Overview: Introducing DTrace
- Summary

Multithread Architecture

- Objectives
- Lesson Agenda
- Common Terminology
- Lesson Agenda
- Multiprocessor Architectures
- Threads in the Solaris 9 and Solaris 10 OS
- Kernel Threads
- Lesson Agenda
- Process Structures
- Fields from the proc Structure
- Fields from the user Structure
- Fields from the Kernel Lightweight Process Structure (klwp_t)
- Fields from the Kernel Thread Structure (kthread_t)
- Fields from Kernel Thread Structure
- cpu Structure

- Lesson Agenda
- Interrupts
- Interrupt Threads
- Interrupt Threads Priorities
- Lesson Agenda
- Locks
- Mutex Locks
- Adaptive Mutex
- Spin Mutex
- Acquiring a Mutex Lock
- Turnstiles
- Semaphores
- Multiple-Reader, Single-Writer Locks
- Condition Variables Structure
- Sleep Queue Properties
- Practice 3 Overview: Multithread Architecture
- Summary

Hardware Memory Management

- Objectives
- Lesson Agenda
- Main Memory
- Virtual Memory
- Process Address Space
- Memory Terminology
- Lesson Agenda
- System Memory Model
- Lesson Agenda
- Virtual-to-Physical Address Translation
- Lesson Agenda
- x86 – 32-Bit MMU
- Page Table Entry
- Page Table Entry (PTE)
- x86 with Physical Addressing Extension (PAE)
- AMD 64-Bit MMU
- Translation Lookaside Buffer (TLB)
- Large Page Sizes
- Spitfire Memory Management Unit (SFMMU)
- Translation Storage Buffer (TSB) Properties
- Translation Storage Buffer
- Table Translation Entry
- ctx Structure
- hme_blks Block
- Lesson Agenda
- Cache

- Cache Compared to Memory
- Cache Hit Rate
- Defining Types of Caches
- Virtual Address Cache
- Physical Address Cache
- Cache Aliasing on a Virtual Cache
- Direct-Mapped Cache
- Set-Associative Cache
- Set-Associative Cache Properties
- Harvard and Unified Caches
- Write-Through and Write-Back Cache
- Cache Snooping
- I/O Cache
- Lesson Agenda
- Hardware Address Translation Layer
- Practice 4 Overview: Hardware Memory Management
- Summary

Software Memory Management

- Objectives
- Lesson Agenda
- SunOS VM1: Features
- Process Address Space
- VM1 Virtual Memory System Layers
- VM2 Virtual Memory System Layers
- What Has Not Changed in Phase 1
- mmap(2) System Call
- System Calls and Services
- madvise(3C) Routine
- NUMA Locality
- Latency Topology: Example
- Ladder Topology: Example
- Address Space Layer
- as Structure
- Physical Pages
- page Structure
- sf_hment Structure
- memseg Structure
- Lesson Agenda
- Virtual Memory Segment Drivers
- seg Structure
- seg_ops Structure
- File I/O Shared-Mapping Segment Driver
- Device Segment Driver (segdev)
- Kernel Memory Segment Driver (segkmem)

- vnode Segment Driver (segvn)
- segvn_data Structure
- vpage Structure
- Anonymous Memory
- AVL Trees
- Mapping Structures
- Process Memory Data Structures
- segkp Driver
- segkp Driver Strategy
- Kernel Physical Mapping Segment Driver (segkpm)
- Practice 5 Overview: Software Memory Management
- Summary

VM2

- Objectives
- Lesson Agenda
- Motivation for VM2
- What Is in VM2 Phase 1?
- Lesson Agenda
- What Has Not Changed in Phase 1?
- VM1 Virtual Memory System Layers
- VM2 Virtual Memory System Layers
- VM2 Phase 2 and Phase 3
- VM2: The Big Picture
- VM2 Update 1
- Proposed Process Mappings
- Lesson Agenda
- Criteria for Memory Selection: mnodes
- mnode
- mnode MDB: Example
- Lesson Agenda
- Criteria for Memory Selection: Tiles
- Tiles
- Tile Sizes
- System Tiles: Example
- Tile Data Structures
- Example in mdb
- Criteria for Memory Selection: tilelets and tilechunks
- Tilelets
- System Tilelets: Example
- Example in mdb
- tilechunk_t
- Tilechunks: Example
- System Tilechunks: Example
- tilechunk_t mdb: Example

- Physical Address of Tilechunk Maps
- ::pachunk – mdb Example
- tileset_t
- ::tileset mdb Example
- Criteria for Memory Selection: Kernel Cage
- VM2
- Kernel Cage in VM2
- Lesson Agenda
- Criteria for Memory Selection: Typed Page Credits
- Typed Page Credits
- Capture
- Bounds Predictor
- Criteria for Memory Selection: memgrp
- memgrp
- memgrp – MDB Example
- System Structure
- System Structure: mdb Example
- Criteria for Memory Selection
- Page Size Codes
- ::size: mdb Example
- Page Allocation Credits
- crd Structure
- crd: mdb Examples
- Wallet
- Fed
- Fed: mdb Example
- Breadline
- Breadline: mdb Examples
- Soupline
- FLR and TCM
- SAC
- SAC: mdb Example
- FLB
- FLB: mdb Examples
- End-to-End Credit Auditing: mdb Example
- Global Page Size Statistics: mdb Examples
- Reverse Map Entry
- RM: Implementation
- RM: Important Flags
- RM: mdb Examples
- RMG
- RMG: mdb Examples
- Sparse Data Structures
- Sparse Data Structures: mdb Examples
- Sparse Data Structures
- Criteria for Memory Selection: Review

- Lesson Agenda
- VM2 Allocation
- Allocation Credits
- Wallet Types
- Freelist Buckets
- Allocation: Glue Interfaces
- Allocation: Locality
- Allocate Credits: Allocation Parameters
- Allocation: Allocate crd_ts
- Allocation: Breadlines
- Allocation: Allocate crd_ts
- Allocation: Allocate rm_ts
- Allocate rm_ts: flr_iterate_slot()
- Allocation: Allocate rm_ts
- Allocate RMs: flr_iterate_tile()
- Allocation: Allocate page_ts
- Lesson Agenda
- Predictor
- Predictor Data Types
- Predictor Threads
- Predictor Components
- Predictor Sampler
- Predictor Analyzer
- Predictor States
- Predictor Action Engine
- Predictor mdb dcmds
- VM2 Structures
- Lesson Agenda
- Procedure for a VM2 Quick Check
- Practice 6 Overview: VM2
- Summary

Paging and Swapping

- Objectives
- Lesson Agenda
- Paging: Overview
- Paging In
- Page Replacement
- Parts of the Page Daemon
- Clock Algorithm
- Defaults for SunOS 5.5.1 Through SunOS 5.8
- Paging Parameters: Updates
- schedpaging() Routine
- pageout()Routine
- pageout_scanner() Routine

- checkpage(pp, whichhand) Routine
- Lesson Agenda
- Swapper
- Swapper (sched.c)Operation
- not_swappable() Macro
- Desperate Swapper Memory Conditions
- sched() Routine, Part 1
- sched() Routine, Part 2
- CL_SWAPOUT() Macro and Routines
- CL_SWAPOUT() Routines
- CL_SWAPIN() Macro and Routines
- rt_swapin()and fx_swapin() Routines
- Lesson Agenda
- Virtual Address Lookup
- Practice 7 Overview: Paging and Swapping
- Summary

The swapfs File System

- Objectives
- Lesson Agenda
- Problems with Anonymous Memory in SunOS 4.x
- Anonymous Memory in SunOS 5.x
- Lesson Agenda
- Swap Management Structures
- anoninfo Structure
- swapinfo Structures
- anon_map Structure
- anon_hdr Structure
- anon Structure in Sun OS 5.x
- Lesson Agenda
- Swap Space Management
- Swap Area: Example
- Lesson Agenda
- Advantages of swapfs File Systems
- Practice 8 Overview: The swapfs File System
- Summary

Scheduling

- Objectives
- Lesson Agenda
- Scheduling Features
- Real-Time Scheduling
- Dispatch Latency
- Pre-Emptible Kernel
- Interactive Scheduling Class

- Fixed Priority Scheduling Class (FX)
- Fair Share Scheduling Class (FSS)
- System Duty Cycle (SDC) Scheduling Class
- SDC Scheduling Class
- Lesson Agenda
- Dispatch Priorities
- dispq_t Structure
- State Diagram
- Class Array
- Lesson Agenda
- Process Structures
- tsproc_t Structure (View 1)
- tsproc_t Structure (View 2)
- rtproc_t Structure
- classfunc Structure
- Lesson Agenda
- Kernel Mode Priority Assignment
- Timesharing Dispatch Parameter Table (View 1)
- Timesharing Dispatch Parameter Table (View 2)
- Timesharing/Interactive Dispatch Parameter Table
- Real-Time Dispatch Parameter Table
- priocntl(1) Command
- dispadmin(1) Command
- Calculating a Thread’s Priority
- Lesson Agenda
- Kernel Scheduling-Related Variables
- Kernel Scheduling-Related Functions
- ts_tick() Routine
- ts_tick() Routine (View 2)
- rt_tick() Routine
- preempt() Routine
- CL_PREEMPT()
- setbackdq(kthread_t *tp)
- ts_preempt(tspp) Routine
- rt_preempt() Routine
- setfrontdq() and setbackdq() Routines
- disp() and swtch()
- disp() Routine
- dispgetwork()
- CMT
- Lesson Agenda
- Priority Inversion
- Callout Queue Processing
- Bounded Priority Inversion
- Unbounded Priority Inversion
- Priority Inheritance

- Blocking Chains
- turnstile_t Structure
- Practice 9 Overview: Scheduling
- Summary

Process Lifetime

- Objectives
- Lesson Agenda
- Process Creation Routines
- Process Structures
- Process Creation System Calls
- Fork Return values
- Fork Extensions
- posix_spawn(3C)Routine
- cfork()
- cfork() Routine
- getproc() and pid_allocate() Routines
- pid_allocate() Routine
- getproc() Routine (View 1)
- getproc() Routine (View 2)
- Process Structures After getproc()
- cfork() and as_dup() Routines
- Return to the cfork() Routine
- Process Structures After as_dup()
- forklwp()and lwp_create() Routines
- thread_create() Routine
- Process Structures After thread_create() and forklwp()
- cfork() and CL_FORKRET() Routines
- Return to cfork()
- Lesson Agenda
- exec Routines
- exec_common() Routine
- execsw Array
- gexec() Routine
- Executable and Linking Format (ELF)
- Elf32_Phdr Program Header Table
- Elf64_Phdr Program Header Table
- Process Segments
- elfexec() Routine
- elfexec() Routine (View 1)
- Auxiliary Vector
- elfexec() Routine (View 2)
- mapelfexec() Routine
- execmap() Routine
- The Initial Process Stack

- proc_exit(why, what) Routine
- waitid (idtype, id, ip, options) Routine
- waitid(idtype, id, ip, options) Routine
- waitid() Routine (View 1)
- waitid() Routine (View 2)
- Practice 10 Overview: Process Lifetime
- Summary

Signals

- Objectives
- Lesson Agenda
- Solaris 10 OS Signals
- Kernel Signal Bitmasks
- Kernel Signal Bitmask: k_sigset_t
- Interrupt and Trap Signals
- Lesson Agenda
- Signal Delivery
- Signal Actions
- Signal Actions: Ignoring
- Signal Actions: Holding
- Kernel-Signal-Related Variables
- Signal Mask Routines
- Assigning a Signal Disposition
- sigaction(sig, *actp, *oactp) System Call
- sigaction(sig, *actp, *oactp): System Call
- sigtoproc(p, t, sig) Routine
- issig(why) Routine
- issig_forreal() Routine
- fsig() Routine
- psig() Routine
- sendsig() Routine
- Calling the Handler
- Practice 11 Overview: Signals
- Summary
- A Appendix A
- A-2
- sysdc_update() A-3
- sysdc_update_pri() A-4