```
**********************************************************
  113266 Fri May  8 18:05:08 2015
new/usr/src/uts/common/vm/seg_dev.c
use NULL dump segop as a shorthand for no-op
Instead of forcing every segment driver to implement a dummy function that
does nothing, handle NULL dump segop function pointer as a no-op shorthand.
**********************************************************
```

```
/*
 * CDDL HEADER START
 *
 * The contents of this file are subject to the terms of the
 * Common Development and Distribution License (the "License").
 * You may not use this file except in compliance with the License.
 *
 * You can obtain a copy of the license at usr/src/OPENSOLARIS.LICENSE
 * or http://www.opensolaris.org/os/licensing.
 * See the License for the specific language governing permissions
 * and limitations under the License.
 *
 * When distributing Covered Code, include this CDDL HEADER in each
 * file and include the License file at usr/src/OPENSOLARIS.LICENSE.
 * If applicable, add the following below this CDDL HEADER, with the
 * fields enclosed by brackets "[]" replaced with your own identifying
 * information: Portions Copyright [yyyy] [name of copyright owner]
 *
 * CDDL HEADER END
 */

/*
 * Copyright 2010 Sun Microsystems, Inc.  All rights reserved.
 * Use is subject to license terms.
 */

/*      Copyright (c) 1984, 1986, 1987, 1988, 1989 AT&T */
/*        All Rights Reserved   */

/*
 * University Copyright- Copyright (c) 1982, 1986, 1988
 * The Regents of the University of California
 * All Rights Reserved
 *
 * University Acknowledgment- Portions of this document are derived from
 * software developed by the University of California, Berkeley, and its
 * contributors.
 */

/*
 * VM - segment of a mapped device.
 *
 * This segment driver is used when mapping character special devices.
 */

#include <sys/types.h>
#include <sys/t_lock.h>
#include <sys/sysmacros.h>
#include <sys/vtrace.h>
#include <sys/systm.h>
#include <sys/vmsystm.h>
#include <sys/mman.h>
#include <sys/errno.h>
#include <sys/kmem.h>
#include <sys/cmn_err.h>
#include <sys/vnode.h>
#include <sys/proc.h>
#include <sys/conf.h>
#include <sys/debug.h>
```



```
#include <sys/ddidevmap.h>
#include <sys/ddi_implfuncs.h>
#include <sys/lgrp.h>

#include <vm/page.h>
#include <vm/hat.h>
#include <vm/as.h>
#include <vm/seg.h>
#include <vm/seg_dev.h>
#include <vm/seg_kp.h>
#include <vm/seg_kmem.h>
#include <vm/vpage.h>

#include <sys/sunddi.h>
#include <sys/esunddi.h>
#include <sys/fs/snode.h>


#if DEBUG
int segdev_debug;
#define DEBUGF(level, args) { if (segdev_debug >= (level)) cmn_err args; }
#else
#define DEBUGF(level, args)
#endif

/* Default timeout for devmap context management */
#define CTX_TIMEOUT_VALUE 0

#define HOLD_DHP_LOCK(dhp)  if (dhp->dh_flags & DEVMAP_ALLOW_REMAP) \
                        { mutex_enter(&dhp->dh_lock); }

#define RELE_DHP_LOCK(dhp) if (dhp->dh_flags & DEVMAP_ALLOW_REMAP) \
                        { mutex_exit(&dhp->dh_lock); }

#define round_down_p2(a, s)     ((a) & ~((s) - 1))
#define round_up_p2(a, s)       (((a) + (s) - 1) & ~((s) - 1))

/*
 * VA_PA_ALIGNED checks to see if both VA and PA are on pgsize boundary
 * VA_PA_PGSIZE_ALIGNED check to see if VA is aligned with PA w.r.t. pgsize
 */
#define VA_PA_ALIGNED(uvaddr, paddr, pgsize)            \
        (((uvaddr | paddr) & (pgsize - 1)) == 0)
#define VA_PA_PGSIZE_ALIGNED(uvaddr, paddr, pgsize)     \
        (((uvaddr ^ paddr) & (pgsize - 1)) == 0)

#define vpgtob(n)       ((n) * sizeof (struct vpage))   /* For brevity */

#define VTOCVP(vp)      (VTOS(vp)->s_commonvp)  /* we "know" it's an snode */

static struct devmap_ctx *devmapctx_list = NULL;
static struct devmap_softlock *devmap_slist = NULL;

/*
 * mutex, vnode and page for the page of zeros we use for the trash mappings.
 * One trash page is allocated on the first ddi_umem_setup call that uses it
 * XXX Eventually, we may want to combine this with what segnf does when all
 * hat layers implement HAT_NOFAULT.
 *
 * The trash page is used when the backing store for a userland mapping is
 * removed but the application semantics do not take kindly to a SIGBUS.
 * In that scenario, the applications pages are mapped to some dummy page
 * which returns garbage on read and writes go into a common place.
 * (Perfect for NO_FAULT semantics)
 * The device driver is responsible to communicating to the app with some
 * other mechanism that such remapping has happened and the app should take
```

```
126  * corrective action.
127  * We can also use an anonymous memory page as there is no requirement to
128  * keep the page locked, however this complicates the fault code. RFE.
129  */
130 static struct vnode trashvp;
131 static struct page *trashpp;

133 /* Non-pageable kernel memory is allocated from the umem_np_arena. */
134 static vmem_t *umem_np_arena;

136 /* Set the cookie to a value we know will never be a valid umem_cookie */
137 #define DEVMAP_DEVMEM_COOKIE    ((ddi_umem_cookie_t)0x1)

139 /*
140  * Macros to check if type of devmap handle
141  */
142 #define cookie_is_devmem(c)     \
143         ((c) == (struct ddi_umem_cookie *)DEVMAP_DEVMEM_COOKIE)

145 #define cookie_is_pmem(c)       \
146         ((c) == (struct ddi_umem_cookie *)DEVMAP_PMEM_COOKIE)

148 #define cookie_is_kpmem(c)      (!cookie_is_devmem(c) && !cookie_is_pmem(c) &&\
149         ((c)->type == KMEM_PAGEABLE))

151 #define dhp_is_devmem(dhp)      \
152         (cookie_is_devmem((struct ddi_umem_cookie *)((dhp)->dh_cookie)))

154 #define dhp_is_pmem(dhp)        \
155         (cookie_is_pmem((struct ddi_umem_cookie *)((dhp)->dh_cookie)))

157 #define dhp_is_kpmem(dhp)       \
158         (cookie_is_kpmem((struct ddi_umem_cookie *)((dhp)->dh_cookie)))

160 /*
161  * Private seg op routines.
162  */
163 static int      segdev_dup(struct seg *, struct seg *);
164 static int      segdev_unmap(struct seg *, caddr_t, size_t);
165 static void     segdev_free(struct seg *);
166 static faultcode_t segdev_fault(struct hat *, struct seg *, caddr_t, size_t,
167                     enum fault_type, enum seg_rw);
168 static faultcode_t segdev_faulta(struct seg *, caddr_t);
169 static int      segdev_setprot(struct seg *, caddr_t, size_t, uint_t);
170 static int      segdev_checkprot(struct seg *, caddr_t, size_t, uint_t);
171 static void     segdev_badop(void);
172 static int      segdev_sync(struct seg *, caddr_t, size_t, int, uint_t);
173 static size_t   segdev_incore(struct seg *, caddr_t, size_t, char *);
174 static int      segdev_lockop(struct seg *, caddr_t, size_t, int, int,
175                     ulong_t *, size_t);
176 static int      segdev_getprot(struct seg *, caddr_t, size_t, uint_t *);
177 static u_offset_t       segdev_getoffset(struct seg *, caddr_t);
178 static int      segdev_gettype(struct seg *, caddr_t);
179 static int      segdev_getvp(struct seg *, caddr_t, struct vnode **);
180 static int      segdev_advise(struct seg *, caddr_t, size_t, uint_t);
181 static void     segdev_dump(struct seg *);
181 static int      segdev_pagelock(struct seg *, caddr_t, size_t,
182                     struct page ***, enum lock_type, enum seg_rw);
183 static int      segdev_getmemid(struct seg *, caddr_t, memid_t *);

185 /*
186  * XXX  this struct is used by rootnex_map_fault to identify
187  *      the segment it has been passed. So if you make it
188  *      "static" you'll need to fix rootnex_map_fault.
189  */
190 const struct seg_ops segdev_ops = {
```



```
191         .dup            = segdev_dup,
192         .unmap          = segdev_unmap,
193         .free           = segdev_free,
194         .fault          = segdev_fault,
195         .faulta         = segdev_faulta,
196         .setprot        = segdev_setprot,
197         .checkprot      = segdev_checkprot,
198         .kluster        = (int (*)())segdev_badop,
199         .sync           = segdev_sync,
200         .incore         = segdev_incore,
201         .lockop         = segdev_lockop,
202         .getprot        = segdev_getprot,
203         .getoffset      = segdev_getoffset,
204         .gettype        = segdev_gettype,
205         .getvp          = segdev_getvp,
206         .advise         = segdev_advise,
208         .dump           = segdev_dump,
207         .pagelock       = segdev_pagelock,
208         .getmemid       = segdev_getmemid,
209 };
```

*unchanged_portion_omitted_*

```
2375 /*
2376  * segdev pages are not dumped, so we just return
2377  */
2378 /*ARGSUSED*/
2379 static void
2380 segdev_dump(struct seg *seg)
2381 {}

2373 /*
2374  * ddi_segmap_setup:    Used by drivers who wish specify mapping attributes
2375  *                      for a segment.  Called from a drivers segmap(9E)
2376  *                      routine.
2377  */
2378 /*ARGSUSED*/
2379 int
2380 ddi_segmap_setup(dev_t dev, off_t offset, struct as *as, caddr_t *addrp,
2381     off_t len, uint_t prot, uint_t maxprot, uint_t flags, cred_t *cred,
2382     ddi_device_acc_attr_t *accattrp, uint_t rnumber)
2383 {
2384         struct segdev_crargs dev_a;
2385         int (*mapfunc)(dev_t dev, off_t off, int prot);
2386         uint_t hat_attr;
2387         pfn_t pfn;
2388         int     error, i;

2390         TRACE_0(TR_FAC_DEVMAP, TR_DEVMAP_SEGMAP_SETUP,
2391             "ddi_segmap_setup:start");

2393         if ((mapfunc = devopsp[getmajor(dev)]->devo_cb_ops->cb_mmap) == nodev)
2394                 return (ENODEV);

2396         /*
2397          * Character devices that support the d_mmap
2398          * interface can only be mmap'ed shared.
2399          */
2400         if ((flags & MAP_TYPE) != MAP_SHARED)
2401                 return (EINVAL);

2403         /*
2404          * Check that this region is indeed mappable on this platform.
2405          * Use the mapping function.
2406          */
2407         if (ddi_device_mapping_check(dev, accattrp, rnumber, &hat_attr) == -1)
2408                 return (ENXIO);
```

```
2410         /*
2411          * Check to ensure that the entire range is
2412          * legal and we are not trying to map in
2413          * more than the device will let us.
2414          */
2415         for (i = 0; i < len; i += PAGESIZE) {
2416                 if (i == 0) {
2417                         /*
2418                          * Save the pfn at offset here. This pfn will be
2419                          * used later to get user address.
2420                          */
2421                         if ((pfn = (pfn_t)cdev_mmap(mapfunc, dev, offset,
2422                             maxprot)) == PFN_INVALID)
2423                                 return (ENXIO);
2424                 } else {
2425                         if (cdev_mmap(mapfunc, dev, offset + i, maxprot) ==
2426                             PFN_INVALID)
2427                                 return (ENXIO);
2428                 }
2429         }

2431         as_rangelock(as);
2432         /* Pick an address w/o worrying about any vac alignment constraints. */
2433         error = choose_addr(as, addrp, len, ptob(pfn), ADDR_NOVACALIGN, flags);
2434         if (error != 0) {
2435                 as_rangeunlock(as);
2436                 return (error);
2437         }

2439         dev_a.mapfunc = mapfunc;
2440         dev_a.dev = dev;
2441         dev_a.offset = (offset_t)offset;
2442         dev_a.type = flags & MAP_TYPE;
2443         dev_a.prot = (uchar_t)prot;
2444         dev_a.maxprot = (uchar_t)maxprot;
2445         dev_a.hat_attr = hat_attr;
2446         dev_a.hat_flags = 0;
2447         dev_a.devmap_data = NULL;

2449         error = as_map(as, *addrp, len, segdev_create, &dev_a);
2450         as_rangeunlock(as);
2451         return (error);

2453 }
```

_____***unchanged_portion_omitted_***

```
**********************************************************
    9136 Fri May  8 18:05:08 2015
new/usr/src/uts/common/vm/seg_kpm.c
use NULL dump segop as a shorthand for no-op
Instead of forcing every segment driver to implement a dummy function that
does nothing, handle NULL dump segop function pointer as a no-op shorthand.
**********************************************************
```

```
/*
 * CDDL HEADER START
 *
 * The contents of this file are subject to the terms of the
 * Common Development and Distribution License, Version 1.0 only
 * (the "License").  You may not use this file except in compliance
 * with the License.
 *
 * You can obtain a copy of the license at usr/src/OPENSOLARIS.LICENSE
 * or http://www.opensolaris.org/os/licensing.
 * See the License for the specific language governing permissions
 * and limitations under the License.
 *
 * When distributing Covered Code, include this CDDL HEADER in each
 * file and include the License file at usr/src/OPENSOLARIS.LICENSE.
 * If applicable, add the following below this CDDL HEADER, with the
 * fields enclosed by brackets "[]" replaced with your own identifying
 * information: Portions Copyright [yyyy] [name of copyright owner]
 *
 * CDDL HEADER END
 */
/*
 * Copyright 2006 Sun Microsystems, Inc.  All rights reserved.
 * Use is subject to license terms.
 */

/*
 * Kernel Physical Mapping (kpm) segment driver (segkpm).
 *
 * This driver delivers along with the hat_kpm* interfaces an alternative
 * mechanism for kernel mappings within the 64-bit Solaris operating system,
 * which allows the mapping of all physical memory into the kernel address
 * space at once. This is feasible in 64 bit kernels, e.g. for Ultrasparc II
 * and beyond processors, since the available VA range is much larger than
 * possible physical memory. Momentarily all physical memory is supported,
 * that is represented by the list of memory segments (memsegs).
 *
 * Segkpm mappings have also very low overhead and large pages are used
 * (when possible) to minimize the TLB and TSB footprint. It is also
 * extentable for other than Sparc architectures (e.g. AMD64). Main
 * advantage is the avoidance of the TLB-shootdown X-calls, which are
 * normally needed when a kernel (global) mapping has to be removed.
 *
 * First example of a kernel facility that uses the segkpm mapping scheme
 * is seg_map, where it is used as an alternative to hat_memload().
 * See also hat layer for more information about the hat_kpm* routines.
 * The kpm facilty can be turned off at boot time (e.g. /etc/system).
 */

#include <sys/types.h>
#include <sys/param.h>
#include <sys/sysmacros.h>
#include <sys/systm.h>
#include <sys/vnode.h>
#include <sys/cmn_err.h>
#include <sys/debug.h>
#include <sys/thread.h>
#include <sys/cpuvar.h>
#include <sys/bitmap.h>
```



```
#include <sys/atomic.h>
#include <sys/lgrp.h>

#include <vm/seg_kmem.h>
#include <vm/seg_kpm.h>
#include <vm/hat.h>
#include <vm/as.h>
#include <vm/seg.h>
#include <vm/page.h>

/*
 * Global kpm controls.
 * See also platform and mmu specific controls.
 *
 * kpm_enable -- global on/off switch for segkpm.
 * . Set by default on 64bit platforms that have kpm support.
 * . Will be disabled from platform layer if not supported.
 * . Can be disabled via /etc/system.
 *
 * kpm_smallpages -- use only regular/system pagesize for kpm mappings.
 * . Can be useful for critical debugging of kpm clients.
 * . Set to zero by default for platforms that support kpm large pages.
 *   The use of kpm large pages reduces the footprint of kpm meta data
 *   and has all the other advantages of using large pages (e.g TLB
 *   miss reduction).
 * . Set by default for platforms that don't support kpm large pages or
 *   where large pages cannot be used for other reasons (e.g. there are
 *   only few full associative TLB entries available for large pages).
 *
 * segmap_kpm -- separate on/off switch for segmap using segkpm:
 * . Set by default.
 * . Will be disabled when kpm_enable is zero.
 * . Will be disabled when MAXBSIZE != PAGESIZE.
 * . Can be disabled via /etc/system.
 *
 */
int kpm_enable = 1;
int kpm_smallpages = 0;
int segmap_kpm = 1;

/*
 * Private seg op routines.
 */
faultcode_t segkpm_fault(struct hat *hat, struct seg *seg, caddr_t addr,
                        size_t len, enum fault_type type, enum seg_rw rw);
static void     segkpm_dump(struct seg *);
static int      segkpm_pagelock(struct seg *seg, caddr_t addr, size_t len,
                        struct page ***page, enum lock_type type,
                        enum seg_rw rw);

static const struct seg_ops segkpm_ops = {
        .fault          = segkpm_fault,
        .dump           = segkpm_dump,
        .pagelock       = segkpm_pagelock,
//#ifndef SEGKPM_SUPPORT
#if 0
#error FIXME: define nop
        .dup            = nop,
        .unmap          = nop,
        .free           = nop,
        .faulta         = nop,
        .setprot        = nop,
        .checkprot      = nop,
        .kluster        = nop,
        .sync           = nop,
        .incore         = nop,
```

```
124         .lockop         = nop,
125         .getprot        = nop,
126         .getoffset      = nop,
127         .gettype        = nop,
128         .getvp          = nop,
129         .advise         = nop,
130         .getpolicy      = nop,
131 #endif
132 };
```

unchanged_portion_omitted_

```
310 #endif  /* SEGKPM_SUPPORT */

312 /* ARGSUSED */
313 static int
314 segkpm_pagelock(struct seg *seg, caddr_t addr, size_t len,
315         struct page ***page, enum lock_type type, enum seg_rw rw)
316 {
317         return (ENOTSUP);
320 }

322 /*
323  * segkpm pages are not dumped, so we just return
324  */
325 /*ARGSUSED*/
326 static void
327 segkpm_dump(struct seg *seg)
328 {
318 }
```

unchanged_portion_omitted_

```
**********************************************************
   82163 Fri May  8 18:05:09 2015
new/usr/src/uts/common/vm/seg_spt.c
use NULL dump segop as a shorthand for no-op
Instead of forcing every segment driver to implement a dummy function that
does nothing, handle NULL dump segop function pointer as a no-op shorthand.
**********************************************************
______unchanged_portion_omitted_

  85 static int segspt_shmdup(struct seg *seg, struct seg *newseg);
  86 static int segspt_shmunmap(struct seg *seg, caddr_t raddr, size_t ssize);
  87 static void segspt_shmfree(struct seg *seg);
  88 static faultcode_t segspt_shmfault(struct hat *hat, struct seg *seg,
  89                 caddr_t addr, size_t len, enum fault_type type, enum seg_rw rw);
  90 static faultcode_t segspt_shmfaulta(struct seg *seg, caddr_t addr);
  91 static int segspt_shmsetprot(register struct seg *seg, register caddr_t addr,
  92                         register size_t len, register uint_t prot);
  93 static int segspt_shmcheckprot(struct seg *seg, caddr_t addr, size_t size,
  94                         uint_t prot);
  95 static int      segspt_shmkluster(struct seg *seg, caddr_t addr, ssize_t delta);
  96 static size_t segspt_shmincore(struct seg *seg, caddr_t addr, size_t len,
  97                         register char *vec);
  98 static int segspt_shmsync(struct seg *seg, register caddr_t addr, size_t len,
  99                         int attr, uint_t flags);
 100 static int segspt_shmlockop(struct seg *seg, caddr_t addr, size_t len,
 101                         int attr, int op, ulong_t *lockmap, size_t pos);
 102 static int segspt_shmgetprot(struct seg *seg, caddr_t addr, size_t len,
 103                         uint_t *protv);
 104 static u_offset_t segspt_shmgetoffset(struct seg *seg, caddr_t addr);
 105 static int segspt_shmgettype(struct seg *seg, caddr_t addr);
 106 static int segspt_shmgetvp(struct seg *seg, caddr_t addr, struct vnode **vpp);
 107 static int segspt_shmadvise(struct seg *seg, caddr_t addr, size_t len,
 108                         uint_t behav);
 109 static void segspt_shmdump(struct seg *seg);
 109 static int segspt_shmpagelock(struct seg *, caddr_t, size_t,
 110                         struct page ***, enum lock_type, enum seg_rw);
 111 static int segspt_shmgetmemid(struct seg *, caddr_t, memid_t *);
 112 static lgrp_mem_policy_info_t *segspt_shmgetpolicy(struct seg *, caddr_t);

 114 const struct seg_ops segspt_shmops = {
 115         .dup            = segspt_shmdup,
 116         .unmap          = segspt_shmunmap,
 117         .free           = segspt_shmfree,
 118         .fault          = segspt_shmfault,
 119         .faulta         = segspt_shmfaulta,
 120         .setprot        = segspt_shmsetprot,
 121         .checkprot      = segspt_shmcheckprot,
 122         .kluster        = segspt_shmkluster,
 123         .sync           = segspt_shmsync,
 124         .incore         = segspt_shmincore,
 125         .lockop         = segspt_shmlockop,
 126         .getprot        = segspt_shmgetprot,
 127         .getoffset      = segspt_shmgetoffset,
 128         .gettype        = segspt_shmgettype,
 129         .getvp          = segspt_shmgetvp,
 130         .advise         = segspt_shmadvise,
 132         .dump           = segspt_shmdump,
 131         .pagelock       = segspt_shmpagelock,
 132         .getmemid       = segspt_shmgetmemid,
 133         .getpolicy      = segspt_shmgetpolicy,
 134 };
______unchanged_portion_omitted_

2794 /*
2795  * We need to wait for pending IO to complete to a DISM segment in order for
2796  * pages to get kicked out of the seg_pcache.  120 seconds should be more
```



```
2797  * than enough time to wait.
2798  */
2799 static clock_t spt_pcache_wait = 120;

2801 /*ARGSUSED*/
2802 static int
2803 segspt_shmadvise(struct seg *seg, caddr_t addr, size_t len, uint_t behav)
2804 {
2805         struct shm_data *shmd = (struct shm_data *)seg->s_data;
2806         struct spt_data *sptd = (struct spt_data *)shmd->shm_sptseg->s_data;
2807         struct anon_map *amp;
2808         pgcnt_t pg_idx;
2809         ushort_t gen;
2810         clock_t end_lbolt;
2811         int writer;
2812         page_t **ppa;

2814         ASSERT(seg->s_as && AS_LOCK_HELD(seg->s_as, &seg->s_as->a_lock));

2816         if (behav == MADV_FREE) {
2817                 if ((sptd->spt_flags & SHM_PAGEABLE) == 0)
2818                         return (0);

2820                 amp = sptd->spt_amp;
2821                 pg_idx = seg_page(seg, addr);

2823                 mutex_enter(&sptd->spt_lock);
2824                 if ((ppa = sptd->spt_ppa) == NULL) {
2825                         mutex_exit(&sptd->spt_lock);
2826                         ANON_LOCK_ENTER(&amp->a_rwlock, RW_READER);
2827                         anon_disclaim(amp, pg_idx, len);
2828                         ANON_LOCK_EXIT(&amp->a_rwlock);
2829                         return (0);
2830                 }

2832                 sptd->spt_flags |= DISM_PPA_CHANGED;
2833                 gen = sptd->spt_gen;

2835                 mutex_exit(&sptd->spt_lock);

2837                 /*
2838                  * Purge all DISM cached pages
2839                  */
2840                 seg_ppurge_wiredpp(ppa);

2842                 /*
2843                  * Drop the AS_LOCK so that other threads can grab it
2844                  * in the as_pageunlock path and hopefully get the segment
2845                  * kicked out of the seg_pcache.  We bump the shm_softlockcnt
2846                  * to keep this segment resident.
2847                  */
2848                 writer = AS_WRITE_HELD(seg->s_as, &seg->s_as->a_lock);
2849                 atomic_inc_ulong((ulong_t *)(&(shmd->shm_softlockcnt)));
2850                 AS_LOCK_EXIT(seg->s_as, &seg->s_as->a_lock);

2852                 mutex_enter(&sptd->spt_lock);

2854                 end_lbolt = ddi_get_lbolt() + (hz * spt_pcache_wait);

2856                 /*
2857                  * Try to wait for pages to get kicked out of the seg_pcache.
2858                  */
2859                 while (sptd->spt_gen == gen &&
2860                     (sptd->spt_flags & DISM_PPA_CHANGED) &&
2861                     ddi_get_lbolt() < end_lbolt) {
2862                         if (!cv_timedwait_sig(&sptd->spt_cv,
```

```
2863                            &sptd->spt_lock, end_lbolt)) {
2864                                break;
2865                        }
2866                }

2868                mutex_exit(&sptd->spt_lock);

2870                /* Regrab the AS_LOCK and release our hold on the segment */
2871                AS_LOCK_ENTER(seg->s_as, &seg->s_as->a_lock,
2872                    writer ? RW_WRITER : RW_READER);
2873                atomic_dec_ulong((ulong_t *)(&(shmd->shm_softlockcnt)));
2874                if (shmd->shm_softlockcnt <= 0) {
2875                        if (AS_ISUNMAPWAIT(seg->s_as)) {
2876                                mutex_enter(&seg->s_as->a_contents);
2877                                if (AS_ISUNMAPWAIT(seg->s_as)) {
2878                                        AS_CLRUNMAPWAIT(seg->s_as);
2879                                        cv_broadcast(&seg->s_as->a_cv);
2880                                }
2881                                mutex_exit(&seg->s_as->a_contents);
2882                        }
2883                }

2885                ANON_LOCK_ENTER(&amp->a_rwlock, RW_READER);
2886                anon_disclaim(amp, pg_idx, len);
2887                ANON_LOCK_EXIT(&amp->a_rwlock);
2888        } else if (lgrp_optimizations() && (behav == MADV_ACCESS_LWP ||
2889            behav == MADV_ACCESS_MANY || behav == MADV_ACCESS_DEFAULT)) {
2890                int                     already_set;
2891                ulong_t                 anon_index;
2892                lgrp_mem_policy_t       policy;
2893                caddr_t                 shm_addr;
2894                size_t                  share_size;
2895                size_t                  size;
2896                struct seg              *sptseg = shmd->shm_sptseg;
2897                caddr_t                 sptseg_addr;

2899                /*
2900                 * Align address and length to page size of underlying segment
2901                 */
2902                share_size = page_get_pagesize(shmd->shm_sptseg->s_szc);
2903                shm_addr = (caddr_t)P2ALIGN((uintptr_t)(addr), share_size);
2904                size = P2ROUNDUP((uintptr_t)(((addr + len) - shm_addr)),
2905                    share_size);

2907                amp = shmd->shm_amp;
2908                anon_index = seg_page(seg, shm_addr);

2910                /*
2911                 * And now we may have to adjust size downward if we have
2912                 * exceeded the realsize of the segment or initial anon
2913                 * allocations.
2914                 */
2915                sptseg_addr = sptseg->s_base + ptob(anon_index);
2916                if ((sptseg_addr + size) >
2917                    (sptseg->s_base + sptd->spt_realsize))
2918                        size = (sptseg->s_base + sptd->spt_realsize) -
2919                            sptseg_addr;

2921                /*
2922                 * Set memory allocation policy for this segment
2923                 */
2924                policy = lgrp_madv_to_policy(behav, len, MAP_SHARED);
2925                already_set = lgrp_shm_policy_set(policy, amp, anon_index,
2926                    NULL, 0, len);

2928                /*
```



```
2929                 * If random memory allocation policy set already,
2930                 * don't bother reapplying it.
2931                 */
2932                if (already_set && !LGRP_MEM_POLICY_REAPPLICABLE(policy))
2933                        return (0);

2935                /*
2936                 * Mark any existing pages in the given range for
2937                 * migration, flushing the I/O page cache, and using
2938                 * underlying segment to calculate anon index and get
2939                 * anonmap and vnode pointer from
2940                 */
2941                if (shmd->shm_softlockcnt > 0)
2942                        segspt_purge(seg);

2944                page_mark_migrate(seg, shm_addr, size, amp, 0, NULL, 0, 0);
2945        }

2947        return (0);
2950 }

2952 /*ARGSUSED*/
2953 void
2954 segspt_shmdump(struct seg *seg)
2955 {
2956         /* no-op for ISM segment */
2948 }
```

_____unchanged_portion_omitted_

****
55158 Fri May 8 18:05:09 2015
new/usr/src/uts/common/vm/vm_seg.c
use NULL dump segop as a shorthand for no-op
Instead of forcing every segment driver to implement a dummy function that
does nothing, handle NULL dump segop function pointer as a no-op shorthand.
****
_____***unchanged_portion_omitted_***

```
1989 void
1990 segop_dump(struct seg *seg)
1991 {
1992         if (seg->s_ops->dump == NULL)
1993                 return;
1992         VERIFY3P(seg->s_ops->dump, !=, NULL);

1995         seg->s_ops->dump(seg);
1996 }
```

_____***unchanged_portion_omitted_***

```
**********************************************************
   16469 Fri May  8 18:05:09 2015
new/usr/src/uts/i86xpv/vm/seg_mf.c
use NULL dump segop as a shorthand for no-op
Instead of forcing every segment driver to implement a dummy function that
does nothing, handle NULL dump segop function pointer as a no-op shorthand.
**********************************************************
______unchanged_portion_omitted_

 472 /*ARGSUSED*/
 473 static void
 474 segmf_dump(struct seg *seg)
 475 {}

 477 /*ARGSUSED*/
 473 static int
 474 segmf_pagelock(struct seg *seg, caddr_t addr, size_t len,
 475     struct page ***ppp, enum lock_type type, enum seg_rw rw)
 476 {
 477         return (ENOTSUP);
 478 }
______unchanged_portion_omitted_

 734 static const struct seg_ops segmf_ops = {
 735         .dup            = segmf_dup,
 736         .unmap          = segmf_unmap,
 737         .free           = segmf_free,
 738         .fault          = segmf_fault,
 739         .faulta         = segmf_faulta,
 740         .setprot        = segmf_setprot,
 741         .checkprot      = segmf_checkprot,
 742         .kluster        = segmf_kluster,
 743         .sync           = segmf_sync,
 744         .incore         = segmf_incore,
 745         .lockop         = segmf_lockop,
 746         .getprot        = segmf_getprot,
 747         .getoffset      = segmf_getoffset,
 748         .gettype        = segmf_gettype,
 749         .getvp          = segmf_getvp,
 750         .advise         = segmf_advise,
 756         .dump           = segmf_dump,
 751         .pagelock       = segmf_pagelock,
 752         .getmemid       = segmf_getmemid,
 753 };
______unchanged_portion_omitted_
```

```
**********************************************************
   11466 Fri May  8 18:05:09 2015
new/usr/src/uts/sparc/v9/vm/seg_nf.c
use NULL dump segop as a shorthand for no-op
Instead of forcing every segment driver to implement a dummy function that
does nothing, handle NULL dump segop function pointer as a no-op shorthand.
**********************************************************
```

```
/*
 * CDDL HEADER START
 *
 * The contents of this file are subject to the terms of the
 * Common Development and Distribution License (the "License").
 * You may not use this file except in compliance with the License.
 *
 * You can obtain a copy of the license at usr/src/OPENSOLARIS.LICENSE
 * or http://www.opensolaris.org/os/licensing.
 * See the License for the specific language governing permissions
 * and limitations under the License.
 *
 * When distributing Covered Code, include this CDDL HEADER in each
 * file and include the License file at usr/src/OPENSOLARIS.LICENSE.
 * If applicable, add the following below this CDDL HEADER, with the
 * fields enclosed by brackets "[]" replaced with your own identifying
 * information: Portions Copyright [yyyy] [name of copyright owner]
 *
 * CDDL HEADER END
 */
/*
 * Copyright 2006 Sun Microsystems, Inc.  All rights reserved.
 * Use is subject to license terms.
 */

/* Copyright (c) 1983, 1984, 1985, 1986, 1987, 1988, 1989 AT&T */
/* All Rights Reserved */

/*
 * Portions of this source code were derived from Berkeley 4.3 BSD
 * under license from the Regents of the University of California.
 */

/*
 * VM - segment for non-faulting loads.
 */

#include <sys/types.h>
#include <sys/t_lock.h>
#include <sys/param.h>
#include <sys/mman.h>
#include <sys/errno.h>
#include <sys/kmem.h>
#include <sys/cmn_err.h>
#include <sys/vnode.h>
#include <sys/proc.h>
#include <sys/conf.h>
#include <sys/debug.h>
#include <sys/archsystm.h>
#include <sys/lgrp.h>

#include <vm/page.h>
#include <vm/hat.h>
#include <vm/as.h>
#include <vm/seg.h>
#include <vm/vpage.h>

/*
 * Private seg op routines.
```



```
 */
static int      segnf_dup(struct seg *seg, struct seg *newseg);
static int      segnf_unmap(struct seg *seg, caddr_t addr, size_t len);
static void     segnf_free(struct seg *seg);
static faultcode_t segnf_nomap(void);
static int      segnf_setprot(struct seg *seg, caddr_t addr,
                    size_t len, uint_t prot);
static int      segnf_checkprot(struct seg *seg, caddr_t addr,
                    size_t len, uint_t prot);
static int      segnf_nop(void);
static int      segnf_getprot(struct seg *seg, caddr_t addr,
                    size_t len, uint_t *protv);
static u_offset_t segnf_getoffset(struct seg *seg, caddr_t addr);
static int      segnf_gettype(struct seg *seg, caddr_t addr);
static int      segnf_getvp(struct seg *seg, caddr_t addr, struct vnode **vpp);
static void     segnf_dump(struct seg *seg);
static int      segnf_pagelock(struct seg *seg, caddr_t addr, size_t len,
                    struct page ***ppp, enum lock_type type, enum seg_rw rw);


const struct seg_ops segnf_ops = {
        .dup            = segnf_dup,
        .unmap          = segnf_unmap,
        .free           = segnf_free,
        .fault          = (faultcode_t (*)(struct hat *, struct seg *, caddr_t,
            size_t, enum fault_type, enum seg_rw))segnf_nomap,
        .faulta         = (faultcode_t (*)(struct seg *, caddr_t)) segnf_nomap,
        .setprot        = segnf_setprot,
        .checkprot      = segnf_checkprot,
        .sync           = (int (*)(struct seg *, caddr_t, size_t, int, uint_t))
                segnf_nop,
        .incore         = (size_t (*)(struct seg *, caddr_t, size_t, char *))
                segnf_nop,
        .lockop         = (int (*)(struct seg *, caddr_t, size_t, int, int,
            ulong_t *, size_t))segnf_nop,
        .getprot        = segnf_getprot,
        .getoffset      = segnf_getoffset,
        .gettype        = segnf_gettype,
        .getvp          = segnf_getvp,
        .advise         = (int (*)(struct seg *, caddr_t, size_t, uint_t))
                segnf_nop,
        .dump           = segnf_dump,
        .pagelock       = segnf_pagelock,
};
______unchanged_portion_omitted_


/*
 * segnf pages are not dumped, so we just return
 */
/* ARGSUSED */
static void
segnf_dump(struct seg *seg)
{}

/*ARGSUSED*/
static int
segnf_pagelock(struct seg *seg, caddr_t addr, size_t len,
    struct page ***ppp, enum lock_type type, enum seg_rw rw)
{
        return (ENOTSUP);
}
______unchanged_portion_omitted_
```