

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Troubleshooting a SUSE® Linux Enterprise 10 System

Leo Hordijk

Overview

- Oh no?!
- Messages during boot
- Boot parameters
- It's all in the logs
- Collecting more information
- Common Hardware issues
- Memory handling
- Filesystem messages
- Start scripts

Oh no?!

Or what can go wrong

Types of Issues

- The system doesn't install
- The system does install, but doesn't come up
- The system seems to lock up after a while
- It actually locks up... or even crashes
- Applications do not start up
- Applications crash
- Expected better performance...
- Everything works ok, but what does that message mean?

Installation Workflow

- Syslinux or pxeboot loads the installation kernel and **initrd** and starts the kernel
- The kernel mounts the **initrd** and starts **linuxrc** as the init process
- **linuxrc** sets up devices required to load the root filesystem and loads it to a RAM disk
- The root system contains YaST, which will request additional information, prepare disk volumes and install packages
- **linuxrc** starts YaST and detaches the **initrd** with a **pivot_root**

Linuxrc booting up

```
N SUSE Linux Enterprise Server
2007 09/10 09:00:00

md: md driver 0.90.3 MAX_MD_DEVS=256, MD_SB_DISKS=27
md: bitmap version 4.39
NET: Registered protocol family 2
IP route cache hash table entries: 4096 (order: 2, 16384 bytes)
TCP established hash table entries: 16384 (order: 4, 65536 bytes)
TCP bind hash table entries: 16384 (order: 4, 65536 bytes)
TCP: Hash tables configured (established 16384 bind 16384)
TCP reno registered
NET: Registered protocol family 1
Using IPI Shortcut mode
ACPI wakeup devices:
  USB
ACPI: (supports S0 S1 S4 S5)
Freeing unused kernel memory: 164k freed
Moving into tmpfs... done.
>>> SUSE Linux Enterprise Server 10 installation program v2.0.67 (c) 1996-2007 SUSE Linux Pro
ducts GmbH <<<
Starting udev ...
... udev running
Starting hardware detection...
Activating usb devices... done
Searching for info file...
Loading Installation System (93480 kB) - 100%
starting hald... ok
starting syslogd (logging to /dev/tty4)... ok
starting klogd... ok
starting yast...
Probing connected terminal...

Initializing virtual console...

Found a Linux console terminal on /dev/console (95 columns x 33 lines).
```

Installation Workflow (cont'd)

- After installation of the packages, YaST starts the installed system with a reboot
- On the first start of the installed system, YaST will be called again, asks some further configuration questions and finishes setup
- Afterwards, the system changes to the default runlevel and is ready to use

```
#
# Let YaST2 finish its installation, if you installed with YaST2
#
if test -f /var/lib/YaST2/runme_at_boot ; then
  HOSTTYPE=$(uname -m)
  splashtrigger "YaST"
  exec 0<> $REDIRECT 1>&0 2>&0
  # if yast2 failed, this ensures proper system setup
  #ulimit -c unlimited
  touch /var/lib/YaST2/run_suseconfig
  if test -x /usr/lib/YaST2/startup/YaST2.Second-Stage; then
    /usr/lib/YaST2/startup/YaST2.Second-Stage
  else
    # oops, yast2 not installed
    rm -f /var/lib/YaST2/runme_at_boot
  fi
fi
# run SuSEconfig (with args) if needed
if test -f /var/lib/YaST2/run_suseconfig ; then
```

312,1

/etc/init.d/boot script

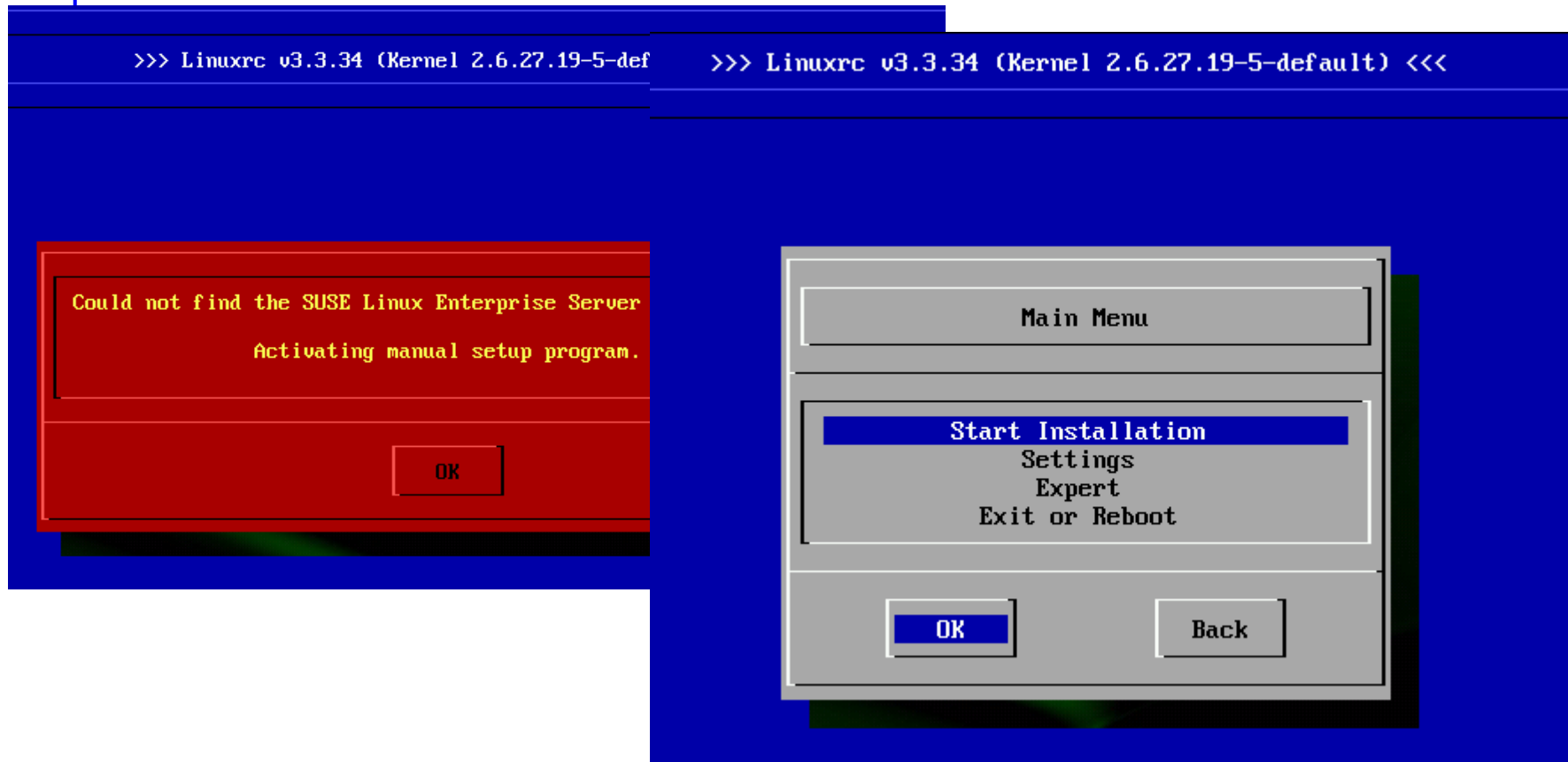
Installation Doesn't Work?

- Syslinux or bootloader doesn't come up
 - Check DVD, drive or BIOS settings
- Syslinux starts, but cannot load kernel or initrd
 - Check installation media (wrongly created?)
- Kernel starts, but hangs at splash screen
 - Press Esc and check messages
- Esc does not work -- still hangs at splash screen
 - Reboot, start with textmode(F3 or text=1) and safemode options
 - Check screen-messages to where it gets stuck

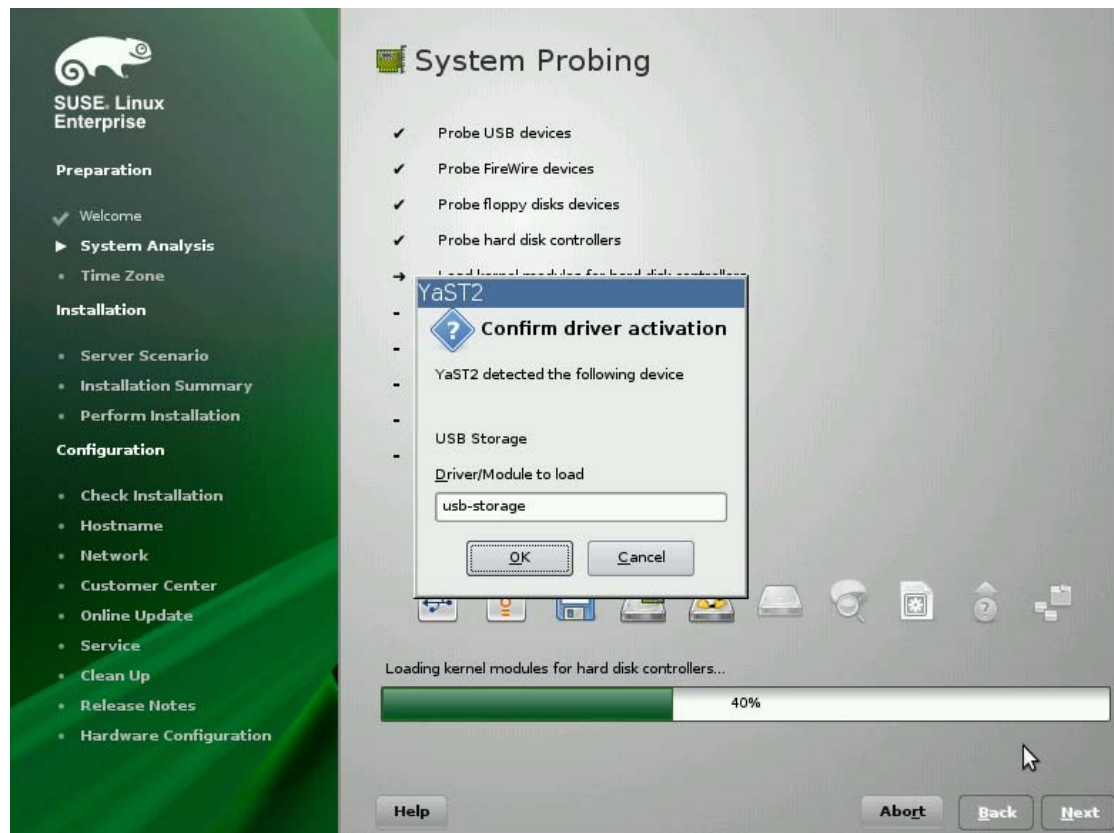
Installation Doesn't Work? (cont'd)

- linuxrc starts, but hangs
 - Try safe mode boot options
- linuxrc starts, but falls back to manual setup
 - Normal for mainframe -- go through manual setup
 - It can't find installation system. Check installation source

Linuxrc falls to manual mode

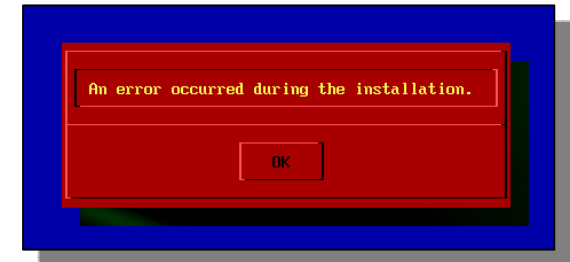


Manual Install



Installation Doesn't Work? (cont'd)

- YaST2 starts, but eventually crashes and falls back to linuxrc
 - Check if disks have correct partition label (Powerpc)
 - Check if filesystems have been created
 - Check if enough RAM (90% of all service requests)
 - Check installation source (completeness)
- Manual partitioning?
 - Use StartShell=1 parameter
 - Use Ctrl-Alt-F2
- Bootloader problems?
 - Boot into linuxrc on reboot by specifying "manual=1"
 - Specify to boot installed system from correct partition



Installation Doesn't Work? (cont'd)

- YaST2 still crashes
 - Check `/var/log/YaST2/y2log`
- There is a shell on text console 2 or 9 (SLES11 - 2,5,6,9)
- On zSeries, there is an sshd process running for login via network during installation
 - force using `ssh=1`
 - connect using `ssh -X ip-address`

```
SUSE Linux Enterprise Server 11 Installation
- there are shells running on consoles 2, 5, 6, 9
- use 'extend' to load extensions (remove with 'extend -r'); extensions are:
  o bind, gdb, sax2
- network setup: run, e.g. 'dhcpcd eth0'
- sshd: run 'rcsshd start' (don't forget to set a password with 'passwd')
/ # _
```

After Installation...

- runlevel says “unknown” runlevel
 - May happen with vnc / ssh / remote X11 installation
- Means YaST2 didn't run after installation
 - vnc installation: start vncviewer
 - X11 installation: start remote X11 server
 - SSH installation:



Novellcongres.nl

13 mei 2009

De Reehorst Ede

Boot Messages

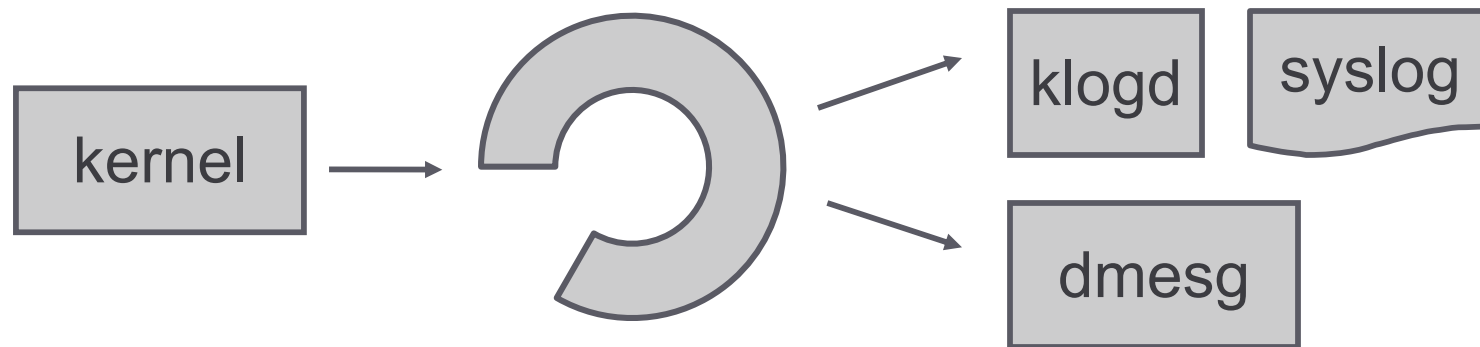
Kernel Boot Messages

- The Linux kernel prints a lot of status messages during startup
- While the amount might be confusing at first, these messages are a valuable resource
- Some messages “smell” like error messages, but are in fact just harmless status notices or warnings
- The kernel sometimes finds flaws of the hardware design or BIOS and complains about them, but works around the limitations

```
ata1.00: ata1: dev 0 multi count 16
ata1.00: configured for UDMA/133
scsil : sata_sis
ata2: SATA link down (SStatus 0 SControl 300)
ATA: abnormal status 0x7F on port 0xB007
  Vendor: ATA          Model: WDC WD3200AAKS-0  Rev: 12.0
  Type:   Direct-Access          ANSI SCSI
revision: 05
SCSI device sda: 625142448 512-byte hdwr sectors (320073 MB)
sda: Write Protect is off
sda: Mode Sense: 00 3a 00 00
SCSI device sda: drive cache: write back
```

Kernel Boot Messages (cont'd)

- Kernel messages are logged by the `klogd` daemon, which feeds them to `syslogd`
- Previous kernel messages can be read with `dmesg`, which reads the `/proc/kmsg` ring buffer



Some Important Kernel Boot Messages

```
Linux version 2.6.16.27-0.6-smp (geeko@buildhost) (gcc
version 4.1.0 (SUSE Linux)) #1 SMP Wed Dec 13 09:34:50
UTC 2006
```

```
BIOS-provided physical RAM map:
```

```
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
```

```
..
```

```
0MB HIGHMEM available.
```

```
384MB LOWMEM available.
```

```
found SMP MP-table at 000f70e0
```

```
On node 0 totalpages: 98304
```

```
DMA zone: 4096 pages, LIFO batch:0
```

```
DMA32 zone: 0 pages, LIFO batch:0
```

```
Normal zone: 94208 pages, LIFO batch:31
```

```
HighMem zone: 0 pages, LIFO batch:0
```

```
DMI present.
```

```
Using APIC driver default
```

```
..
```

```
Kernel command line: root=/dev/sda2 vga=0x333
```

```
resume=/dev/sda1 splash=silent
```

Some Important Kernel Boot Messages (cont'd)

```
Detected 3007.353 MHz processor.
```

```
Using tsc for high-res timesource
```

```
..
```

```
Memory: 382340k/393216k available (1611k kernel code,  
10180k reserved, 730k data, 188k init, 0k highmem)
```

```
Calibrating delay using timer specific routine.. 3202.94
```

```
BogoMIPS (lpj=6405897)
```

```
..
```

```
CPU0: Intel(R) Xeon(TM) CPU 1.60GHz stepping 08
```

```
Total of 1 processors activated (3202.94 BogoMIPS).
```

```
ENABLING IO-APIC IRQs
```

```
..
```

```
PCI: Using ACPI for IRQ routing
```

Some Important Kernel Boot Messages (cont'd)

```
PIIX4: IDE controller at PCI slot 0000:00:07.1
PIIX4: chipset revision 1
PIIX4: not 100% native mode: will probe irqs later
idel: BM-DMA at 0x1058-0x105f, BIOS settings: hdc:DMA,
hdd:pio
Probing IDE interface idel1...
hdc: VMware Virtual IDE CDROM Drive, ATAPI CD/DVD-ROM drive
idel at 0x170-0x177,0x376 on irq 15
..
scsi0 : BusLogic BT-958
Vendor: VMware, Model: VMware Virtual S Rev: 1.0
Type: Direct-Access SCSI revision: 02
SCSI device sda: 12582912 512-byte hdwr sectors (6442 MB)
.. (ignore SCSI errors here) ..
sda: assuming drive cache: write through
sda: sda1 sda2
sd 0:0:0:0: Attached scsi disk sda
```

Kernel Oops

They look like this:

```
Oops: 0002 [#1] SMP
last sysfs file: /class/net/sit0/address
Modules linked in: xt_pkttype ipt_LOG xt_limit ...
CPU: 0
EIP: 0061:[<c901d1cf>] Not tainted
EFLAGS: 00210082 (2.6.16.14-6-xenpae #1)
EIP is at network_tx_buf_gc+0xcf/0x270 [xennet]
eax: 00000077 ebx: 00000077 ecx: 0000007e edx:
00000000
esi: 00000076 edi: c7f48380 ebp: 01dabb76 esp:
c646f968
ds: 007b es: 007b ss: 0069
Process httpd2-prefork (pid: 20004, threadinfo=c646e000
task=c7a34270)
Stack: c1de5b80 c646e000 c7f48000 01dabb76 01dabb78
00000000 c7f4840c c7f4838
..
```

Kernel Oops (cont'd)

Call Trace:

```
[<c901e560>] netif_int+0x30/0x140 [xenet]  
[<c0144eb8>] handle_IRQ_event+0x38/0xd0  
[<c0144fed>] __do_IRQ+0x9d/0x110  
[<c0106977>] do_IRQ+0x37/0x70  
[<c0117409>] __wake_up_common+0x39/0x60
```

..

```
Code: 00 00 8d 87 e4 08 00 00 e8 4f 62 23 f7 c7 84 9f e8  
08 00 00 00 00 00 00 8b
```

```
Kernel panic - not syncing: Fatal exception in interrupt
```

Kernel Oops (cont'd)

- Not every kernel stack trace is necessarily a kernel problem
- Three kinds of causes for kernel Oops:
 - Invalid pointer dereferences
 - BUG() assertions
 - Jump to invalid address
- Kernel Oops within an interrupt handler will halt the kernel
- Check if there is a newer maintenance kernel available
 - In many cases installing a newer kernel release solves the issue

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Boot Parameters



Boot Parameters

- If the kernel starts, but the system eventually locks up, try the Failsafe selection of the boot loader
- Failsafe kernel parameters:
 - `ide=nodma` only use PIO modes for IDE disks - slow!
 - `edd=off` BIOS Extended Disk Driver services off
 - `apm=off` don't use APM
 - `acpi=off` don't use ACPI
 - `noapic` fall back to XT interrupt controller
 - `nosmp` turn off SMP support...
 - `maxcpus=0` ...and don't even look at other CPUs

Beware of Anagrams

- ACPI is the Advanced Configuration and Power Interface, an API of the PC's BIOS.
Originally designed as a power management interface, it now serves as a hardware configuration information hub provided by the BIOS
- APIC is the Advanced Programmable Interrupt Controller, a part of the mainboard's chipset.
Additionally to the original XT PC PIC (i8259A), it can distribute interrupts from more than 16 sources to any of the CPUs of an SMP system
- To add to the confusion, an ACPI table contains the APIC configuration data

ACPI parameters

- `acpi=off`
turns off any ACPI support. This might result in bad interrupt routing, wrong memory configuration, faulty PCI bus setup or not recognizing all CPUs
- `acpi=oldboot`
changes the handling of ACPI tables. Some BIOSes still do not initialize their ACPI tables correctly, resulting in broken interrupt routing or faulty PCI bus setup. This option takes care about that

ACPI parameters (cont'd)

- `acpi=force`
turns on ACPI support even though the BIOS has been blacklisted in the kernel
ACPI: BIOS listed in blacklist, disabling ACPI support
- `acpi=ht`
uses ACPI only to enable hyperthreading
- `acpi=noirq`
uses ACPI, but not for interrupt routing
- `pci=noacpi`
disables ACPI for interrupt routing and PCI scanning

How To Find IRQ Routing Issues

- Warning and error messages regarding irq routing in `/var/log/messages` or `/var/log/boot.msg`

```
ACPI: Unable to set IRQ for PCI Interrupt Link [LNKE] (likely buggy
ACPI BIOS).
Aborting ACPI-based IRQ routing. Try pci=noacpi or acpi=off
```

- Look at `/proc/interrupts` if IRQs are distributed well via available interrupt lines and all CPUs
- Edge-triggered IRQs shouldn't be shared

```
localadmin@node1:/proc> cat interrupts
          CPU0          CPU1
0: 494968601          0      IO-APIC-edge  timer
1:          10          0      IO-APIC-edge  i8042
8:           2          0      IO-APIC-edge  rtc
9:           0          0      IO-APIC-level acpi
12:          124          0      IO-APIC-edge  i8042
14:          5448 26594623      IO-APIC-edge  ide0
15:           111          164      IO-APIC-edge  ide1
169:          441 307364186      IO-APIC-level  libata, eth4
185:           46          0      IO-APIC-level  SiS SI7012
193:           0          0      IO-APIC-level  ehci_hcd:usb1
201:           0          0      IO-APIC-level  ohci_hcd:usb2
209:           0          0      IO-APIC-level  ohci_hcd:usb3
217:           0          0      IO-APIC-level  ohci_hcd:usb4
NMI:           0          0
LOC: 494963530 494963530
ERR:           0
MIS:           0
```

System Clock Issues

- Some BIOSes do not configure the chipset correctly, resulting in some clock source not working properly or even in double timer ticks
- Sometimes `clock=pit` helps, but reduces accuracy
- Interrupt timeout on timer tick while kernel boot. Doesn't hurt (unless kernel locks up), kernel will try another source
- Some chipsets require `noapic`
- In rare cases, dynamic CPU frequency results in wrong system times. Turn off power management

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Collecting More Information

Important Logfiles

- `/var/log/messages`
The central log file. Almost everything goes here
- `/var/log/boot.msg`
The kernel boot and start script messages
- `/var/log/mcelog`
Machine Check Exceptions
- `/var/log/YaST2/y2log`
YaST2 log files

Enable SysRq

- `echo 1 >/proc/sys/kernel/sysrq`
To enable SysRQ handling on the fly
- Set `ENABLE_SYSRQ="yes"` in
`/etc/sysconfig/sysctl` to activate it on boot
- Run `klogconsole -r0 -18`
to enable verbose kernel logging to console

Trigger SysRq

- To trigger SysRq:
 - press `ALT-SysRq-<letter>` on keyboard
 - press `BREAK <letter>` on serial console

```
echo '<letter>' >/proc/sysrq-trigger
```

on command line
- `ALT-SysRQ-h` for help
- `ALT-SysRQ-m` for memory status
- `ALT-SysRQ-b` for immediate reboot
- `ALT-SysRQ-t` for process list
- `ALT-SysRQ-u` for unmounting disks
- `ALT-SysRQ-s` for sync (dirty-cache to disk)
- Bijvoorbeeld `Alt-SysRq-s`, `Alt-SysRq-u`, `Alt-SysRq-b`.

Serial Console

- **Problem:** The system appears to be dead, the screen saver has kicked in, no way to tell if there was a kernel oops...
- **Solution:** Use a serial console to log messages
- **Required:** A serial terminal, or a PC with a terminal emulator and a null-modem cable
- **Configuration:**
add `console=ttyS0,9600`
to the kernel boot parameters. Boot the machine and increase kernel logging (klogconsole)

When You Open a Support Call

- What happened?
- What brand and model of the machine?
- Have the boot options mentioned in the YES bulletin been used?
- What is the content of `/var/log/messages` and `/var/log/boot.msg`
- Send result of `hwinfo --log hwinfo.txt`
- Send result of `getsysinfo`
- On System z: result of `dbginfo.sh`

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Hardware Issues

Typical Hardware Issues

- Machine crashes without apparent reason, or kernel Oops shows an EIP outside the kernel memory
 - Check /var/log/mcelog
 - Check RAM with memtest86
 - Check disks and cables
 - Check power supply, grounding, AC

Typical Hardware Issues (cont'd)

- Unstable network link status
 - Check network topology
 - Check cables
 - Check NIC
 - Check switches and routers
 - Check grounding

Typical Hardware Issues (cont'd)

- Disk and multipathing problems
 - Check disks
 - Check storage array configuration
 - Check cables
 - Check switches
- SCSI subsystem errors in `/var/log/messages` about bus resets, non-responding targets or parity errors are almost always hardware errors

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Memory Handling

Memory Handling

- # free

```
                total    used    free shared buffers cached
Mem:           1035504 874492 161012      0   48592 480076
-/+ buffers/cache: 345824 689680
Swap:          1052216  10116 1042100
#
```

- It is normal that that “free” shows a large amount of cached buffers
- Linux does not swap-in swapped memory unless needed, thus a large amount of swapped-out memory despite plenty of cache memory is possible

Memory Handling (cont'd)

- A huge amount of cache with only a small amount of free memory is not a problem per se
- If the machine is constantly swapping in this situation, it might not be able to write the dirty cache to disk in time, or in fact has it in use
- Either the machine does not have enough I/O bandwidth, or the I/O scheduler is not suitable for this kind of load. Try `elevator=as` in the latter case
- On machines with low I/O latency, try setting `pdflush` parameters to more synchronous writes
 - `/proc/sys/vm/nr_pdflush_threads`
 - <http://www.westnet.com/~gsmith/content/linux-pdflush.htm>

Memory Handling (cont'd)

- For cache values that don't decrease when needed, locked SHM pages may be the cause. Take a look at `ipcs -m` and `df /dev/shm/`
- kernel: `__alloc_pages: 0-order allocation failed (gfp=0xf0/0)`
is not as bad as it sounds if there are only few of them. The Kernel cannot allocate memory for its own purposes (usually network drivers.)

Out of Memory

- OOM killer terminating processes? Either some process runs amok or the machine does not have enough memory
- YaST dying on installation? Add more memory
- For large third-party packages: implement the memory configuration recommendations of the vendor
- Use the sa tools of the sysstat package to monitor memory and I/O usage

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Filesystem Messages

Broken Filesystem?

- **Problem:** `fsck` complains about a broken file system on every boot, even though the machine was shut down properly
- **Cause:** Quite often this is caused by SATA drives that claim to support tagged queuing or write back on power failure
- **Remedy:** Boot with kernel parameter `barrier=off`
- **Note:** SoftRAID mirroring may also suffer from this

Filesystem Goes Read-Only?

- **Problem:**
EXT3-fs warning (device dm-3): ext3_unlink:
Deleting nonexistent file (1274784), 0
EXT3-fs error (device dm-3):
ext3_free_blocks: bit already cleared for
block 2914565
Aborting journal on device dm-3.
Remounting filesystem read-only
- **Cause:** The Filesystem on dm-3 is already in use by some other machine
- **Solution:** Fix the HA setup: take care that a node can only come up if nothing uses the same resources. E.g., use stonith

Fixing ReiserFS

- **Problem:** fsck complains about broken ReiserFS at boot
- **Analysis:**
 - Does it still occur with `barrier=off`?
 - Is the LVM group / MD setup / EVMS volume there?
 - Is the volume writeable?
 - Is the disk undamaged?
- **Repair:**
 1. `reiserfsck --fix-fixable` should do in most cases
 2. `reiserfsck -rebuild-tree` usually fixes the rest
 3. `reiserfsck -rebuild-sb` for the desperate

Novellcongres.nl

13 mei 2009

De Reehorst Ede

Start Scripts

The Mysteriously Moved Start Script

- **Problem:** A start script for a third-party application has been installed in `/etc/init.d/myapp` and manually linked to `/etc/init.d/rc3/S25myapp`. After a while the link gets renamed to `/etc/init.d/rc3.d/S00myapp`
- **Cause:** The script does not have the LSB headers for start/stop scripts. The next `chkconfig` or `insserv` renames it to `S00<name>` as it does not have any dependencies to other scripts defined
- **Solution:** Add the LSB header and install start scripts with `chkconfig` or `insserv`

The Mysteriously Moved Start Script (cont'd)

- This is an example for an LSB header for a start script:

```
#!/bin/sh
#
### BEGIN INIT INFO
# Provides:          updatedns
# Required-Start:    $named
# Should-Start:      $named ntp
# Required-Stop:     $named
# Should-Stop:       $named ntp
# Default-Start:     3 5
# Default-Stop:      0 1 2 6
# Short-Description: updates example.com DNS
# Description:       updates example.com DNS on
request
### END INIT INFO
#
```

- See also: `/etc/init.d/skeleton`, `/etc/insserv.conf` and `insserv(8)`

Niet vergeten!

Kom langs op de stand van Twice in de Verdifoyer en

- Schrijf u in voor de 1 daagse Workshop “Troubleshooten” het vervolg op deze presentatie;
- Beantwoord de prijsvraag en maak kans op

Vragen en antwoorden

