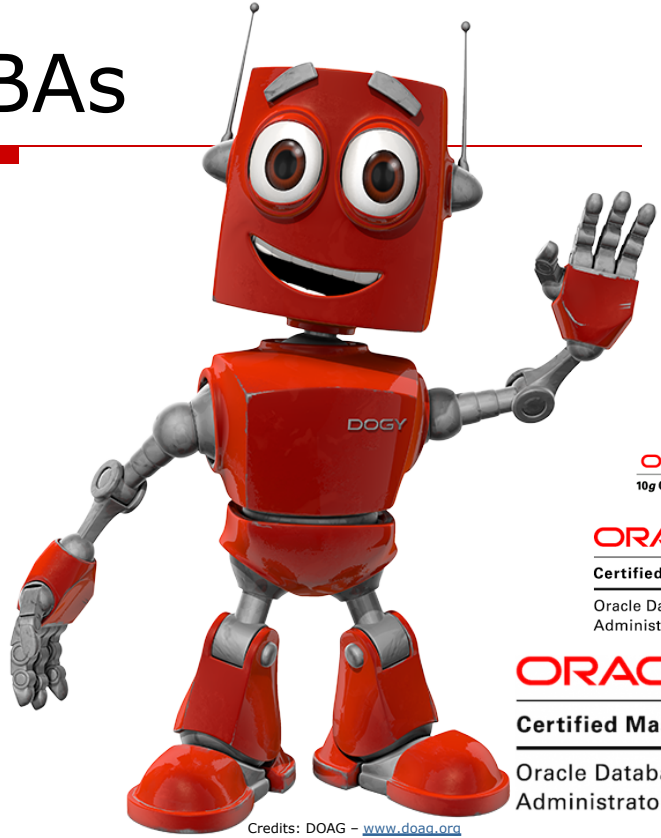
ORA-SOLUTIONS.NET
Mastering Oracle Performance, High Availability, Manageability

Martin Decker

LINUX

KNOW-HOW FÜR DBAs



ORACLE
10g Certified Master

ORACLE
Certified Master
Oracle Database 11g
Administrator

ORACLE
Certified Master
Oracle Database 12c
Administrator

Credits: DOAG – www.doag.org

Agenda

- Vorstellung
- What´s Linux
- Installation / Konfiguration
- NUMA
- Big 4: CPU / Memory / Disk IO / Netzwerk
- (Live)-Demos

whoami

- seit 2003 im Oracle Datenbank Umfeld tätig
- seit 2008 unabhängiger Oracle Consultant in D/A/CH
- Spezialisierung auf:
  - Performance Management (Instance / SQL)
  - Hochverfügbarkeit (MAA, RAC, DataGuard)
  - Manageability (OEM)
  - Unix (Linux, Solaris, HP-UX)
- Oracle Certified Master 10g/11g/12c
- Website & Blog: http://www.ora-solutions.net

What´s Linux?

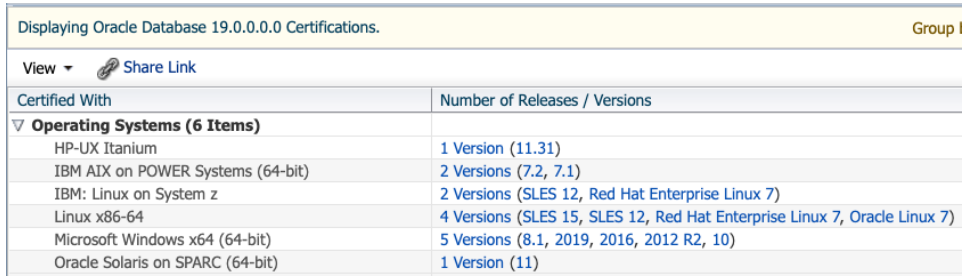
- erste Version des Linux Kernels von Linus Torvalds 1991

  I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones.
- heute unzählige Distributionen*)
- Linux ist frei (GNU GPL) / nur Support ist kostenpflichtig
- "Enterprise Linux"
  - Linux Distribution mit (teilweise) kostenpflichtigen Support
  - Dokumentation
  - nur Enterprise Linux für Oracle Database zertifiziert: OL, SLES, RHEL

*https://de.wikipedia.org/wiki/Liste_von_Linux-Distributionen

What´s Linux?

- Oracle Datenbank wird auf Linux-Plattform entwickelt und anschließend portiert
- Oracle Datenbank-Plattform mit größter Kundenbasis
- Standard-Plattform auch für Engineered Systems (Exadata, ODA)
- zertifiziert für Oracle Database 19c – Achtung RHEL8/OL8 (noch) nicht!



Displaying Oracle Database 19.0.0.0.0 Certifications. Group I

View ▾ Share Link

Certified With	Number of Releases / Versions
Operating Systems (6 Items)	
HP-UX Itanium	1 Version (11.31)
IBM AIX on POWER Systems (64-bit)	2 Versions (7.2, 7.1)
IBM: Linux on System z	2 Versions (SLES 12, Red Hat Enterprise Linux 7)
Linux x86-64	4 Versions (SLES 15, SLES 12, Red Hat Enterprise Linux 7, Oracle Linux 7)
Microsoft Windows x64 (64-bit)	5 Versions (8.1, 2019, 2016, 2012 R2, 10)
Oracle Solaris on SPARC (64-bit)	1 Version (11)

What´s Linux? – Distributionen

- Debian (ca. alle 2 jahre Major Release)
- Fedora Linux (Red-Hat sponsored community linux)
- Enterprise Linux:
  - Ubuntu (basiert auf Debian) von Canonical
  - Red Hat Enterprise Linux *)
    - zB RHEL7 basiert auf Fedora 19, RHEL8 basiert auf Fedora 28, aktuell RHEL8.1
    - Minor Releases alle 6 Monate
    - nur kommerziell erhältlich
  - CentOS Enterprise Linux (Community Enterprise Operating System)
    - basiert auf RHEL / frei verfügbare binärkompatible Linux-Distribution, rebranding
  - Oracle (Enterprise) Linux (früher OEL, nun OL)
    - rebranding und mit RHEL binärkompatible Linux-Distribution
    - Oracle Unbreakable Enterprise Kernel (aktuelle Kernel auch mit älterem OL, zB OL6)
  - Novell SLES (SuSE Linux Enterprise Server)

* https://de.wikipedia.org/wiki/Red_Hat_Enterprise_Linux

What´s Linux?

- Oracle (Enterprise) Linux (früher OEL, nun OL)
  - kostenfrei verfügbar, auch Errata Packages über public-yum Repository
  - kostenpflichtig ist nur der Support
  - bei Premier Support auch Ksplice zum Online Patching des Kernel
  - Support Varianten (Limited: only 2 Sockets per Server)
    - Network Support:
      24x7 Access to Patches/Updates/Security Fixes über ULN (108,-€ p.a./server)
    - Basic Support:
      plus 24x7 service requests (Limited: 444,- € / Normal: 1.080,- € p.a. /server)
    - Permier Support:
      plus KSplice (Limited: 1.260,- € / Normal: 2.064 € p.a. /server)
    - Lizenziert wird Physik – beliebig viele VMs nutzbar

What´s Linux?

- Oracle Unbreakable Enterprise Kernels (UEK)

UEK5: ACFS ab 19.4.0

OL Release	Initial UEK kernel	Initial RHCK kernel	UEK2 supported	UEK3 supported	UEK4 supported	UEK5 supported
OL6U0	kernel-uek-2.6.32-100.28.5	kernel-2.6.32-71	No. OL6U2 required	No. OL6U5 required	No. OL6U8 required	Not Available
OL6U1	kernel-uek-2.6.32-100.34.1	kernel-2.6.32-131.0.15	No. OL6U2 required	No. OL6U5 required	No. OL6U8 required	Not Available
OL6U2	kernel-uek-2.6.32-300.3.1	kernel-2.6.32-220	Yes	No. OL6U5 required	No. OL6U8 required	Not Available
OL6U3	kernel-uek-2.6.39-200.24.1	kernel-2.6.32-279	Yes, default	No. OL6U5 required	No. OL6U8 required	Not Available
OL6U4	kernel-uek-2.6.39-400.17.1	kernel-2.6.32-358	Yes, default	No. OL6U5 required	No. OL6U8 required	Not Available
OL6U5	kernel-uek-3.8.13-16.2.1	kernel-2.6.32-431	Yes	Yes, default	No. OL6U8 required	Not Available
OL6U6	kernel-uek-3.8.13-44.1.1	kernel-2.6.32-504	Yes	Yes, default	No. OL6U8 required	Not Available
OL6U7	kernel-uek-3.8.13-68.3.4	kernel-2.6.32-573	Yes	Yes, default	No. OL6U8 required	Not Available
OL6U8	kernel-uek-4.1.12-37.4.1	kernel-2.6.32-642	Yes	Yes	Yes, default	Not Available
OL6U9	kernel-uek-4.1.12-61.1.28	kernel-2.6.32-696	Yes	Yes	Yes, default	Not Available
OL6U10	kernel-uek-4.1.12-124.16.4	kernel-2.6.32-754	Yes	Yes	Yes, default	Not Available
OL7U0	kernel-uek-3.8.13-35.3.1	kernel-3.10.0-123	Not Available	Yes, default	No. OL7U3 required	No. OL7U5 required
OL7U1	kernel-uek-3.8.13-55.1.6	kernel-3.10.0-229	Not Available	Yes, default	No. OL7U3 required	No. OL7U5 required
OL7U2	kernel-uek-3.8.13-98.7.1	kernel-3.10.0-327	Not Available	Yes, default	No. OL7U3 required	No. OL7U5 required
OL7U3	kernel-uek-4.1.12-61.1.18	kernel-3.10.0-514	Not Available	Yes	Yes, default	No. OL7U5 required
OL7U4	kernel-uek-4.1.12-94.3.9	kernel-3.10.0-693	Not Available	Yes	Yes, default	No. OL7U5 required
OL7U5	kernel-uek-4.1.12-112.16.4	kernel-3.10.0-862	Not Available	Yes	Yes, default	Yes
OL7U6	kernel-uek-4.14.35-1818.3.3	kernel-3.10.0-957	Not Available	Yes	Yes	Yes, default
OL7U7	kernel-uek-4.14.35-1902.3.2	kernel-3.10.0-1062	Not Available	Yes	Yes	Yes, default
OL8U0	Not Available	kernel-4.18.0-80	Not Available	Not Available	Not Available	Not Available

* https://blogs.oracle.com/scoter/oracle-linux-and-unbreakable-enterprise-kernel-uek-releases

Installation / Konfiguration

- Default-RPM Installation (MOS 376183.1)
- Prereq-Packages: yum install <prereq-package>
  - oracle-database-preinstall-19c
  - oracle-database-preinstall-18c
  - oracle-database-server-12cR2-preinstall
- Erstellt/Konfiguriert:
  - User oracle / Groups dba, install, etc.
  - User Limits (number of open files, number of processes, etc.)
  - Disabled Transparent Hugepages (THP*)
  - sysconfig Kernel Parameter
- Erstellt NICHT:
  - start/stop Scripts (init.d, systemd)

* **ALERT: Disable Transparent HugePages on SLES11, RHEL6, RHEL7, OL6, OL7, and UEK2 and above (Doc ID 1557478.1)**

Installation / Konfiguration

- Red Hat Enterprise Linux - Oracle Database Deployment Guide:

https://access.redhat.com/documentation/en-us/reference_architectures/2017/html-single/deploying_oracle_database_12c_release_2_on_red_hat_enterprise_linux_7/index

- **vm.dirty_background_ratio** is the percentage of system memory which when dirty then system can start writing data to the disks.
- **vm.dirty_ratio** is percentage of system memory which when dirty, the process doing writes would block and write out dirty pages to the disks.

Tuned Parameters	balanced	throughput-performance	tuned-profiles-oracle
I/O Elevator	deadline	deadline	deadline
CPU governor	OnDemand	performance	performance
kernel.sched_min_granularity_ns	auto-scaling	10ms	10ms
kernel.sched_wake_up_granularity_ns	3ms	15ms	15ms
disk read-ahead	128 KB	4096 KB	4096 KB
vm.dirty_ratio	20%	40%	80%*
File-system barrier	on	on	on
Transparent HugePages	on	on	off
vm.dirty_background_ratio	10%	10%	3%*
vm.swappiness	60%	10%	1%*

Installation / Konfiguration

tuned

```
$ tuned-adm list
Available profiles:
- balanced                    - General non-specialized tuned profile
- desktop                     - Optimize for the desktop use-case
- hpc-compute                 - Optimize for HPC compute workloads
- latency-performance         - Optimize for deterministic performance at the cost of
                                increased power consumption
- network-latency             - Optimize for deterministic performance at the cost of
                                increased power consumption,
                                focused on low latency network performance
- network-throughput          - Optimize for streaming network throughput, generally
                                 only necessary on older CPUs or 40G+ networks
- powersave                   - Optimize for low power consumption
- throughput-performance      - Broadly applicable tuning that provides excellent
                                performance across a variety of common server workloads
- virtual-guest               - Optimize for running inside a virtual guest
- virtual-host                - Optimize for running KVM guests
Current active profile: throughput-performance
```

Installation / Konfiguration

tuned

```
$ cat /usr/lib/tuned/throughput-performance/tuned.conf  | grep -v ^# | grep -v ^$
[main]
summary=Broadly applicable tuning / excellent performance across variety of common server workloads

[cpu]
governor=performance
energy_perf_bias=performance
min_perf_pct=100

[disk]
readahead=>4096

[sysctl]
kernel.sched_min_granularity_ns = 10000000
kernel.sched_wakeup_granularity_ns = 15000000
vm.dirty_ratio = 40
vm.dirty_background_ratio = 10
vm.swappiness=10
```

CPU Frequency Scaling:
Power Saving vs. Performance

bei sequentiellem Lese-Pattern
liest OS 4096 kB "in advance"

Kernel Parameter *

swappiness: Aggressivität mit der der Kernel bei Memory Pressure Speicherseiten auslagert

* **Oracle Linux 7 - sysctl parameter doesn't take effect after reboot (Doc ID 2195319.1)**

Installation / Konfiguration

- Achtung: CleanUp /var/tmp/.oracle

Oracle Linux 7 and Redhat Linux 7: The socket files in /var/tmp/.oracle Location Get Deleted (Doc ID 2455193.1)

CAUSE

Both Oracle Linux 7 and Redhat Linux 7 have a kernel service systemd-tmpfiles-clean.service that is managed by systemd and deletes the files in the temp location.

The above service removes

1. files/directories in /tmp/ un-accessed more than 10 days(defined in tmp.conf)
2. files/directories in /var/tmp/ un-accessed more than 30 days(defined in tmp.conf)

The "un-accessed" is decided by checking all of atime/mtime/ctime of the file/directory.

SOLUTION

Exclude the socket files from getting deleted by the kernel service systemd-tmpfiles-clean.service

To exclude the socket files in a tmp directory from getting deleted by the tempfile clean service, change the content of /usr/lib/tmpfiles.d/tmp.conf and add

x /tmp/.oracle*
x /var/tmp/.oracle*
x /usr/tmp/.oracle*

The "x" option above indicates to the systemd-tmpfiles-clean.service to exclude files in the listed directory.

Installation / Konfiguration

- Achtung: prelink RPM Package deinstallieren!

Connected to an idle instance, while database instance is running (Process J000 died, see its trace file kkjcre1p: unable to spawn jobq slave process) (Doc ID 1578491.1)

connecting to database says "Connected to an idle instance" while database is running.

```
SQL*Plus: Release 11.2.0.3.0 Production on Thu Aug 22 05:32:03 2013

Copyright (c) 1982, 2011, Oracle.  All rights reserved.

Enter user-name: / as sysdba
Connected to an idle instance.

SQL> exit
```

strace -a -e -f -tt -o connect.out sqlplus "/as sysdba"

```
32259 05:32:10.347860 shmat(7208991, 0x38a0000000, 0) = -1 EINVAL (Invalid argument)  ===>
32259 05:32:10.347925 shmdt(0x60000000) = 0
32259 05:32:10.348002 shmdt(0x160000000) = 0
32259 05:32:10.348296 shmdt(0xc00000000) = 0
```

SOLUTION

The prelink is removed from exadata version 11.2.3.3 release onwards.

For older release,

Remove prelink RPM

rpm -e prelink*

or

1. Remove execute permission chmod -x /usr/sbin/prelink

2. Remove prelink scriptfile from /etc/cron.daily

Why not able to allocate a more SGA than 193G on Linux 64? (Doc ID 1241284.1)
Solution ... by oracle, investigating the issue further, seems that This is caused by the **prelink** command.
Refine to All > Oracle Database Products > Oracle Database Suite > Oracle Database > Oracle Database - Enterprise Edition

PROCESS j000 And m000 Die (Doc ID 790397.1)
Solution - Ensure **prelink** is not run while database is running.The **prelink** will try to arrange shared library mappings in the virtual address space ...
Refine to All > Oracle Cloud > Oracle Infrastructure Cloud > Oracle Cloud at Customer > Oracle Database Exadata Cloud Machine

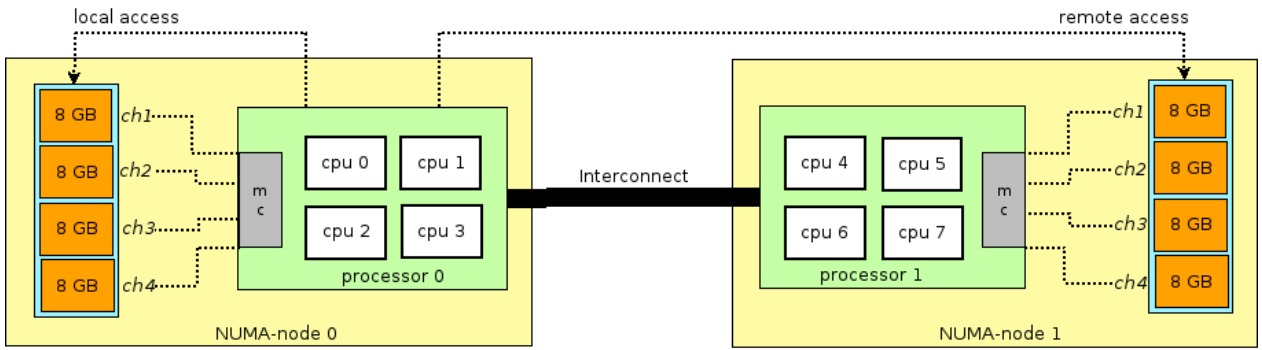
Cannot Restart Oracle After Setting Global Name to NULL (Doc ID 743676.1)
Solution warning: difference appears to be caused by **prelink**, adjusting expectations
Refine to All > Oracle Cloud > Oracle Infrastructure Cloud > Oracle Cloud at Customer > Oracle Database Exadata Cloud Machine

Connected to an idle instance, while database instance is running (Process j000 died, see its trace file kkjcre1p: unable to spawn jobq slave process) (Doc ID 1578491.1)
Cause 18617 root 39 19 2192 1448 420 D 7.7 0.0 0:00.44 /usr/sbin/**prelink** -av -mR -q
Refine to All > Oracle Cloud > Oracle Infrastructure Cloud > Oracle Cloud at Customer > Oracle Database Exadata Cloud Machine

Ora-00600 [18062] During Database Startup (Doc ID 1431334.1)
Solution warning: difference appears to be caused by **prelink**, adjusting expectations
Refine to All > Oracle Cloud > Oracle Infrastructure Cloud > Oracle Cloud at Customer > Oracle Database Exadata Cloud Machine

NUMA – Non Uniform Memory Access

- Systeme mit mehr als 1 CPU Socket
- jeder CPU ist eine Anzahl von (lokalen) Memory Sockets zugewiesen
- direkter Zugriff der CPU auf ihren lokalen Memory
- indirekter Zugriff auf Memory der anderen CPUs über QuickPath Interconnect (QPI)
- NUMA Support in OS per default aktiv




```
 numactl  -H
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11
12 26 27 28 29 30 31 32 33 34 35 36 37 38
node 0 size: 257383 MB
node 0 free: 9013 MB
node 1 cpus: 13 14 15 16 17 18 19 20 21
22 23 24 25 39 40 41 42 43 44 45 46 47
48 49 50 51
node 1 size: 258043 MB
node 1 free: 2229 MB
node distances:
node   0   1
  0:  10  21
  1:  21  10
```


* alert log:
Opening with internal Resource Manager plan where NUMA PG = 2, CPUs = 24

NUMA – Non Uniform Memory Access

- NUMA Support in RDBMS ab 11gR2 default deaktiviert
- bei Engineered Systems (Exadata, ODA) aktiviert
- aktivierbar über init.ora:
  - _enable_NUMA_support=TRUE *
  - _px_numa_support_enabled" = TRUE *
- Optimierung von parallel query und DBWr Operationen durch Präferieren von local memory
- Überlegungen:
  - nur relevant wenn > 1 Socket, je mehr Sockets desto wichtiger
  - wenn Oracle DB NUMA enabled wird, ausgiebig testen über verschiedene Stages
  - bei Virtualisierung: Guest VM wenn möglich in einem NUMA Node halten (Memory, CPUs)

* alert log:
Opening with internal Resource Manager plan where NUMA PG = 2, CPUs = 24

Big 4: CPU / Memory / Disk IO / Network

- CPU:
  - Sockets / Cores / Threads (Hyperthreading)
  - Überlegungen Hardware-Beschaffung bzgl. Oracle Lizenzierung:
    - SE: beliebig viele Cores aber nur max. 2 CPU Sockets, Lizenzkosten pro CPU
    - EE: möglichst wenig Cores, aber dafür maximale Frequenz, Lizenzkosten pro Core * CF (0,5)
  - cat /proc/cpuinfo

Big 4: CPU / Memory / Disk IO / Network

- CPU:
  - lscpu

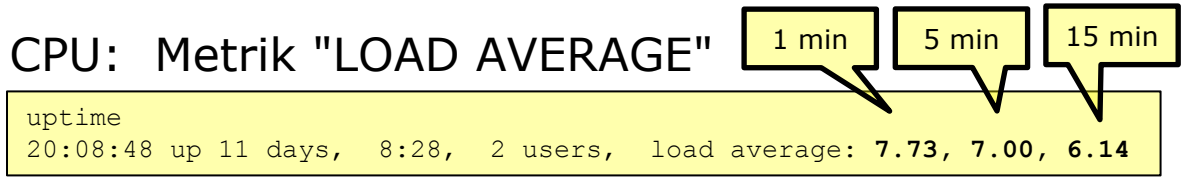
```
[user@host ~]$ lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                32
On-line CPU(s) list:   0-31
Thread(s) per core:    2
Core(s) per socket:    8
Socket(s):             2
NUMA node(s):          2
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 45
Model name:            Intel(R) Xeon(R) CPU E5-2690 0 @ 2.90GHz
Stepping:              7
CPU MHz:               3396.851
CPU max MHz:           3800.0000
CPU min MHz:           1200.0000
BogoMIPS:              5799.97
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              256K
L3 cache:              20480K
NUMA node0 CPU(s):     0-7,16-23
NUMA node1 CPU(s):     8-15,24-31
Flags:                 fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi
mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc arch_perfmon pebs bts rep_good nopl
xtopology nonstop_tsc aperfmperf eagerfpu pni pclmulqdq dtes64 monitor ds_cpl vmx smx est tm2 ssse3 cx16 xtpr
pdcm pcid dca sse4_1 sse4_2 x2apic popcnt tsc_deadline_timer aes xsave avx lahf_lm ida arat epb pln pts dtherm
ibrs stibp ibpb pti tpr_shadow vnmi flexpriority ept vpid xsaveopt
```

2 Sockets / 16 Cores / 32 Threads

2 NUMA Nodes!

Big 4: CPU / Memory / Disk IO / Network

CPU: Metrik "LOAD AVERAGE"



- Durchschnittliche Last für letzte 1, 5 und 15 Min
- Erkennbar, ob Last tendenziell steigt, fällt oder gleich bleibt
- Metrik ist KEIN Indikator bzgl CPU Auslastung
- denn, anders als andere UNIXs enthält die Load Average bei Linux nicht nur
  - runnable processes (running + runnable)
    sondern auch
  - blocked processes (z.B. waiting for i/o)

* http://www.brendangregg.com/blog/2017-08-08/linux-load-averages.html

Big 4: CPU / Memory / Disk IO / Network

- CPU: Falle: "CPU Utilization global vs. pro Core"
  - Benutzer nutzt "vmstat" output um CPU Utilization zu beurteilen
  - Server hat viele CPUs/Cores
  - Obwohl ein Core zu 100% busy ist, ist globale Utilization nur 1/#Anzahl Cores
  - Beispiel:
    - Host mit 32 Cores
    - Oracle Session ist zu 100% auf CPU aktiv
    - d.h. 1 CPU Core ist 100% busy
    - vmstat zeigt CPU Utilization als 1/32 = 3%
    - top zeigt den Oracle Prozess mit max. 100% CPU Utilization an, da Oracle single-threaded
    - bei Java Prozessen oder anderen multi-threaded processes, CPU Utilization angezeigt in top kann 100% steigen

4 Cores verfügbar, global 22% busy

```
top - 22:02:35 up  5:10,  4 users,  load average: 2.57, 1.58, 0.88
Tasks: 527 total,   2 running, 232 sleeping,   0 stopped,   0 zombie
%Cpu(s): 22.4 us,  0.4 sy,  0.0 ni, 77.0 id,  0.1 wa,  0.0 hi,  0.2 si,  0.0 st
KiB Mem : 49179044 total,  7094620 free, 39717428 used,  2366996 buff/cache
KiB Swap:  2093052 total,  2093052 free,        0 used.  8971000 avail Mem
```

1 prozess = 1 core busy

```
  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
29752 oracle    20   0   36.5g 193528 104632 R 100.0  0.4   1:59.45 ora_j003_CAOT194
21077 oracle    20   0   36.4g  98156  91132 S   8.6  0.2   8:40.73 oracleCAOT194 (LOCAL=NO)
```

Big 4: CPU / Memory / Disk IO / Network

Memory: cat /proc/meminfo

```
MemTotal:       528028212 kB
MemFree:        23091108 kB
MemAvailable:   85043064 kB

Buffers:           30048 kB
Cached:         72895288 kB
SwapCached:            4 kB

Active:         45487588 kB
Inactive:       36310540 kB
Active(anon):    9318832 kB
Inactive(anon):   766436 kB
Active(file):   36168756 kB
Inactive(file): 35544104 kB

SwapTotal:      33554428 kB
SwapFree:       33554200 kB
```

MemTotal: Physical Memory: 503 GB sichtbar (512 installed)

MemFree: komplett unbenutzer Memory: 22 GB / vm.min_free_kbytes

MemAvailable: freier (reclaimable) Memory (ohne swap)

Buffers: raw block i/o buffers (eher unwichtig)

Cached: Linux Page Cache (Disk-Cache and Shared Memory)

Active: Recently Used: Active(anon) + Active (file)

Inactive (anon+file): Not recently Used – can be swapped out or reclaimed

Pages swapped = SwapTotal - SwapFree

Big 4: CPU / Memory / Disk IO / Network

- Memory: cat /proc/meminfo

```
...
AnonPages:        9940700 kB
Mapped:            630628 kB

Slab:             1871440 kB
SReclaimable:     1449644 kB
SUnreclaim:        421796 kB
KernelStack:        24640 kB
PageTables:        260056 kB
...
AnonHugePages:          0 kB
...
HugePages_Total:    204800
HugePages_Free:     32512
HugePages_Rsvd:     18564
HugePages_Surp:         0
Hugepagesize:        2048 kB
```

SLAB: Kernel Memory für Anforderung unterschiedlicher Größen

PageTables: Mapping zw. virt / phys Memory Pages pro Prozess stark reduziert, wenn Hugepages im Einsatz sind

AnonHugepages: Transparent Hugepages -> Please disable!

Hugepages:
HugePages_Total: gesamte Anzahl an statisch konfigurierten Hugepages à 2 MB
Hugepages Rsvd: noch nicht allokiert, aber reserviert
Hugepages_Free: noch nicht allokiert (aber evtl reserviert)

Big 4: CPU / Memory / Disk IO / Network

Memory: Aufteilung pro NUMA Node

`cat /sys/devices/system/node/node0/meminfo`

```
Node 0 Active:          23785752 kB
Node 0 Active(anon):     6223124 kB
Node 0 Active(file):    17562628 kB
Node 0 AnonHugePages:          0 kB
Node 0 AnonPages:        6588412 kB
Node 0 Bounce:                 0 kB
Node 0 Dirty:                388 kB
Node 0 FilePages:       35495168 kB
Node 0 HugePages_Free:    1206
Node 0 HugePages_Surp:       0
Node 0 HugePages_Total: 102400
Node 0 Inactive:        17611568 kB
Node 0 Inactive(anon):    343500 kB
Node 0 Inactive(file): 17268068 kB
Node 0 KernelStack:        11200 kB
Node 0 Mapped:             297460 kB
Node 0 MemFree:         10643368 kB
Node 0 MemTotal:       263790136 kB
Node 0 MemUsed:        253146768 kB
Node 0 Mlocked:            281424 kB
Node 0 NFS_Unstable:           0 kB
Node 0 PageTables:          92936 kB
Node 0 Shmem:              201080 kB
Node 0 Slab:              1063828 kB
Node 0 SReclaimable:       864492 kB
Node 0 SUnreclaim:         199336 kB
Node 0 Unevictable:        281420 kB
Node 0 Writeback:              0 kB
Node 0 WritebackTmp:           0 kB
```

`cat /sys/devices/system/node/node1/meminfo`

```
Node 1 Active:          21969348 kB
Node 1 Active(anon):     2982536 kB
Node 1 Active(file):    18986812 kB
Node 1 AnonHugePages:          0 kB
Node 1 AnonPages:        3160724 kB
Node 1 Bounce:                 0 kB
Node 1 Dirty:                 28 kB
Node 1 FilePages:       37803504 kB
Node 1 HugePages_Free:    32756
Node 1 HugePages_Surp:       0
Node 1 HugePages_Total: 102400
Node 1 Inactive:        18618424 kB
Node 1 Inactive(anon):    360392 kB
Node 1 Inactive(file): 18258032 kB
Node 1 KernelStack:        12864 kB
Node 1 Mapped:             328116 kB
Node 1 MemFree:         12266384 kB
Node 1 MemTotal:       264238076 kB
Node 1 MemUsed:        251971692 kB
Node 1 Mlocked:             97856 kB
Node 1 NFS_Unstable:           0 kB
Node 1 PageTables:         150708 kB
Node 1 Shmem:              260868 kB
Node 1 Slab:               820988 kB
Node 1 SReclaimable:       599176 kB
Node 1 SUnreclaim:         221812 kB
Node 1 Unevictable:         97856 kB
Node 1 Writeback:              0 kB
Node 1 WritebackTmp:           0 kB
```

Big 4: CPU / Memory / Disk IO / Network

- Memory: Zones
  - Memory wird in zones aufgesplittet:
    - DMA: (0-16 MB) historisch, früher für Driver nötig
    - DMA32 (16-4096 MB): eingeführt für die Übergangszeit zu 64bit für spezielle Hardware
    - Normal: (>4096MB)
  - Nur Normal Zone weiter aufgesplittet pro NUMA Node (später mehr dazu)
  - Kernel benötigt für verschiedene Operationen zusammenhängenden (contiguous) memory
  - Problem: Fragmentierung (analog zu Oracle DB shared pool)
  - Es kann sein, dass trotz genügend freiem Memory kein "contiguous Chunk" mit der gewünschten Größe frei ist
  - Vermeidung: Kernel führt "compaction" * durch

```
Mon Feb 26 08:53:37 2018
skgxpvfynet: mtype: 61 process 15801 failed because of a resource problem in the OS. The OS has most likely run out of buffers (rval: 4)
Errors in file /u01/app/oracle/diag_p/diag/rdbms/prod/PROD2/trace/PROD2_ora_15801.trc (incident=480004):
ORA-00603: ORACLE server session terminated by fatal error
ORA-27504: IPC error creating OSD context
ORA-27300: OS system dependent operation:sendmsg failed with status: 105
ORA-27301: OS failure message: No buffer space available
ORA-27302: failure occurred at: sskgxpsnd2
Incident details in: /u01/app/oracle/diag_p/diag/rdbms/prod/PROD2/incident/incdir_480004/PROD2_ora_15801_i480004.trc
```

```
cat /proc/meminfo
zzz ***Mon Feb 26 08:53:09 CET 201
MemTotal:       528028424 kB
MemFree:        14593828 kB
MemAvailable:   78305772 kB
Buffers:        28009752 kB
Cached:         46896496 kB
SwapCached:            0 kB
Active:         22627436 kB
Inactive:       66945168 kB
Active(anon):   14315300 kB
Inactive(anon):  2105748 kB
Active(file):    8312136 kB
Inactive(file): 64839420 kB
Unevictable:      363996 kB
Mlocked:          364020 kB
SwapTotal:      33554428 kB
SwapFree:       33554428 kB
Dirty:                404 kB
Writeback:              0 kB
```

* https://savvinov.com/2019/10/14/memory-fragmentation-the-silent-performance-killer/

Big 4: CPU / Memory / Disk IO / Network

Memory: Fragmentation

cat /proc/buddyinfo bzw. /tmp/buddyinfo.sh

```
DMA Memory (0-16MB)
DMA:    2^     0 Order (       4 kB):          0 kB (Chunks:         0)
DMA:    2^     1 Order (       8 kB):          0 kB (Chunks:         0)
DMA:    2^     2 Order (      16 kB):         48 kB (Chunks:         3)
DMA:    2^     3 Order (      32 kB):         64 kB (Chunks:         2)
DMA:    2^     4 Order (      64 kB):          0 kB (Chunks:         0)
DMA:    2^     5 Order (     128 kB):        128 kB (Chunks:         1)
DMA:    2^     6 Order (     256 kB):        256 kB (Chunks:         1)
DMA:    2^     7 Order (     512 kB):          0 kB (Chunks:         0)
DMA:    2^     8 Order (    1024 kB):       1024 kB (Chunks:         1)
DMA:    2^     9 Order (    2048 kB):       2048 kB (Chunks:         1)
DMA:    2^    10 Order (    4096 kB):      12288 kB (Chunks:         3)
                            TOTAL:       15856 kB

DMA32 (16-4096M):
DMA32:  2^     0 Order (       4 kB):        136 kB (Chunks:        34)
DMA32:  2^     1 Order (       8 kB):        152 kB (Chunks:        19)
DMA32:  2^     2 Order (      16 kB):        208 kB (Chunks:        13)
DMA32:  2^     3 Order (      32 kB):        384 kB (Chunks:        12)
DMA32:  2^     4 Order (      64 kB):       1024 kB (Chunks:        16)
DMA32:  2^     5 Order (     128 kB):        512 kB (Chunks:         4)
DMA32:  2^     6 Order (     256 kB):       1792 kB (Chunks:         7)
DMA32:  2^     7 Order (     512 kB):       3584 kB (Chunks:         7)
DMA32:  2^     8 Order (    1024 kB):       3072 kB (Chunks:         3)
DMA32:  2^     9 Order (    2048 kB):       4096 kB (Chunks:         2)
DMA32:  2^    10 Order (    4096 kB):    1007616 kB (Chunks:       246)
                            TOTAL:     1038432 kB

Normal Node 1 (>4096M):
Normal: 2^     0 Order (       4 kB):    1638632 kB (Chunks:    409658)
Normal: 2^     1 Order (       8 kB):    2084000 kB (Chunks:    260500)
Normal: 2^     2 Order (      16 kB):     950448 kB (Chunks:     59403)
Normal: 2^     3 Order (      32 kB):     104672 kB (Chunks:      3271)
Normal: 2^     4 Order (      64 kB):      25600 kB (Chunks:       400)
Normal: 2^     5 Order (     128 kB):      19456 kB (Chunks:       152)
Normal: 2^     6 Order (     256 kB):      13056 kB (Chunks:        51)
Normal: 2^     7 Order (     512 kB):      12800 kB (Chunks:        25)
Normal: 2^     8 Order (    1024 kB):      73728 kB (Chunks:        72)
Normal: 2^     9 Order (    2048 kB):       4096 kB (Chunks:         2)
Normal: 2^    10 Order (    4096 kB):    2375680 kB (Chunks:       580)
                            TOTAL:     8340600 kB

Normal Node 2 (>4096M):
Normal: 2^     0 Order (       4 kB):    1363756 kB (Chunks:    340939)
Normal: 2^     1 Order (       8 kB):    1466760 kB (Chunks:    183345)
Normal: 2^     2 Order (      16 kB):    1847440 kB (Chunks:    115465)
Normal: 2^     3 Order (      32 kB):    1386080 kB (Chunks:     43315)
Normal: 2^     4 Order (      64 kB):    1053184 kB (Chunks:     16456)
Normal: 2^     5 Order (     128 kB):     864768 kB (Chunks:      6756)
Normal: 2^     6 Order (     256 kB):     510720 kB (Chunks:      1995)
Normal: 2^     7 Order (     512 kB):     434688 kB (Chunks:       849)
Normal: 2^     8 Order (    1024 kB):     220160 kB (Chunks:       215)
Normal: 2^     9 Order (    2048 kB):      38912 kB (Chunks:        19)
Normal: 2^    10 Order (    4096 kB):    2265088 kB (Chunks:       553)
                            TOTAL:    19792156 kB
```

cat /proc/pagetypeinfo

```
Page block order: 9
Pages per block:  512

Free pages count per migrate type at order       0      1      2      3      4      5      6      7      8      9     10
Node    0, zone      DMA, type    Unmovable      0      0      3      2      0      1      1      0      1      0      0
Node    0, zone      DMA, type  Reclaimable      0      0      0      0      0      0      0      0      0      0      0
Node    0, zone      DMA, type      Movable      0      0      0      0      0      0      0      0      0      0      3
Node    0, zone      DMA, type      Reserve      0      0      0      0      0      0      0      0      0      1      0
Node    0, zone      DMA, type          CMA      0      0      0      0      0      0      0      0      0      0      0
Node    0, zone      DMA, type      Isolate      0      0      0      0      0      0      0      0      0      0      0
Node    0, zone    DMA32, type    Unmovable     26     11      7      9     14      4      3      2      0      0      1
Node    0, zone    DMA32, type  Reclaimable      1      1      2      0      0      0      0      1      1      1      0
Node    0, zone    DMA32, type      Movable      7      7      4      3      2      0      4      4      2      0    244
Node    0, zone    DMA32, type      Reserve      0      0      0      0      0      0      0      0      0      1      1
Node    0, zone    DMA32, type          CMA      0      0      0      0      0      0      0      0      0      0      0
Node    0, zone    DMA32, type      Isolate      0      0      0      0      0      0      0      0      0      0      0
Node    0, zone   Normal, type    Unmovable   4409   2994   2854    483    242     86     43     17      1      0      0
Node    0, zone   Normal, type  Reclaimable      7      2      7     45     26     56      6      6      4      1      0
Node    0, zone   Normal, type      Movable 427085 271426  63199   3247    151     10      2      2     67      1    579
Node    0, zone   Normal, type      Reserve      0      0      0      0      0      0      0      0      0      0      1
Node    0, zone   Normal, type          CMA      0      0      0      0      0      0      0      0      0      0      0
Node    0, zone   Normal, type      Isolate      0      0      0      0      0      0      0      0      0      0      0

Number of blocks type     Unmovable  Reclaimable      Movable      Reserve          CMA      Isolate
Node 0, zone      DMA             1            0            6            1            0            0
Node 0, zone    DMA32             6            2         1006            2            0            0
Node 0, zone   Normal           512          432       129102            2            0            0
Page block order: 9
Pages per block:  512

Free pages count per migrate type at order       0      1      2      3      4      5      6      7      8      9     10
Node    1, zone   Normal, type    Unmovable  15048   6408   1362    468    120     59     17      3      0      0      1
Node    1, zone   Normal, type  Reclaimable     83      5      3      0      1      1     11      7      2      1      0
Node    1, zone   Normal, type      Movable 346966 180916 114943  43083  16376   6697   1969    839    213     18    547
Node    1, zone   Normal, type      Reserve      0      0      0      0      0      0      0      0      0      0      1
Node    1, zone   Normal, type          CMA      0      0      0      0      0      0      0      0      0      0      4
Node    1, zone   Normal, type      Isolate      0      0      0      0      0      0      0      0      0      0      0

Number of blocks type     Unmovable  Reclaimable      Movable      Reserve          CMA      Isolate
Node 1, zone   Normal           598          298       130166            2            8            0
```

Big 4: CPU / Memory / Disk IO / Network

- Memory: Swapping unbedingt vermeiden!!!!

so (kB swap out/sec): Beleg für Memory-Pressure (erstmal unkritisch)
si (kB swap in/sec): Prozesse benötigen Pages von Swap (kritisch)

erste Zeile enthält historische Werte und sollte ignoriert werden

```
$ vmstat 5
procs -----------memory---------- ---swap-- -----io---- -system-- ----cpu----
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa
 3  0 833704  54824  25196 328672   10    0   343    18  510 1382 96  4  0  0
 6  0 833704  54556  25092 324584    0    0   333    22  504 1180 93  7  0  0
 4  0 833704  51516  25112 320856   33    0   315    19  508 1234 95  5  0  0
 3  0 833704  54836  24984 314404    6    0   223    27  498 1191 95  5  0  0
 3  0 833704  53072  24944 307844    4    0   216    22  518 1375 96  4  0  0
 5  0 833704  53928  24888 304076    6    0   262    18  548 1665 94  6  0  0
 3  4 843964  50192    184  58064   16 2416    16  2464  570 1451 78 22  0  0
 3  7 908244  48756    224  47760  118 13645  149 13664  730 1245 76 16  0  8
 3  2 922064  54280    340  49228 1470  2838 1817  2865  711 1481 88 12  0  0
 4  2 932644  54068    424  52204 1972  2195 2596  2211  678 1388 90 10  0  0
 2  3 944012  56304    492  52292 2986  2591 3063  2615  735 1562 89 11  0  0
 2  4 957304  54604    572  51964 4042  3414 4096  3438  852 1808 88 12  0  0
```

Big 4: CPU / Memory / Disk IO / Network

- Memory: swapping-Analyse mit "dstat –mgsp"

```
------memory-usage----- ---paging-- ----swap--- ---procs---
 used  buff  cach  free|  in   out | used  free|run blk new
61.1G 3140k  396M 1341M|   0     0 |   0    64G|4.0   0 1.0
61.1G 3140k  396M 1341M|   0     0 |   0    64G|2.0   0 2.0
62.1G 2164k  244M  475M|   0     0 |   0    64G|5.0   0 3.0
62.1G 2164k  244M  475M|   0     0 |   0    64G|6.0   0   0
62.1G 2164k  244M  475M|   0     0 |   0    64G|4.0   0 1.0
62.1G 2164k  244M  475M|   0     0 |   0    64G|4.0   0 1.0
62.1G 2164k  244M  475M|   0     0 |   0    64G|4.0   0 2.0
62.1G 2164k  244M  475M|   0     0 |   0    64G|3.0   0 1.0
61.7G    0   145M  997M|2348k 1561M|1562M   62G| 13 6.0  11
61.7G    0   169M  955M|2288k 6196k|1567M   62G| 11 6.0 3.0
61.7G    0   182M  935M|8596k    0 |1560M   62G| 20 2.0 1.0
61.7G    0   183M  896M|  38M    0 |1540M   62G|8.0 8.0 1.0
61.7G    0   186M  866M|  27M    0 |1529M   62G|7.0   0 1.0
61.7G    0   186M  865M| 700k    0 |1528M   62G|8.0   0   0
61.7G    0   187M  861M|4136k    0 |1523M   62G| 10   0 1.0
61.8G    0   187M  837M|  24M    0 |1513M   62G| 13   0 1.0
61.8G    0   187M  833M|3372k    0 |1508M   62G| 17   0 1.0
38.8G    0   187M 23.8G| 904k    0 |1506M   62G|8.0 1.0 1.0
```

procs new:
Anzahl der neu erstellten Prozesse (siehe execsnoop)

run/blk
used swap
paging out 1.5G
paging in

Big 4: CPU / Memory / Disk IO / Network

- Memory: Hugepages aktivieren!
  - gesamter Memory wird in 4 kB Pages (getconf PAGE_SIZE) verwaltet
  - im Kernel-Memory liegen die PageTables: Referenztabellen **für jeden OS Prozess**, für das Mapping zwischen virtual Memory Address und Physical Memory Address
  - für Datenbanken mit großen SGAs (init.ora: sga_target) und vielen Prozessen (init.ora: processes) können diese PageTables sehr groß werden. (mehrere GB)
  - der Memory für diese PageTables steht dann nicht für Benutzer-Prozesse (SGA) oder Caching zur Verfügung und Management der PageTables kostet sys-CPU%.
  - das Hugepages werden nicht auf SWAP ausgelagert
  - KONFIGURATION: (init.ora: memory_target/memory_max_target nicht verwenden!)
    - /etc/sysctl.conf:
      ```
      vm.nr_hugepages = <pages in 2MB Einheiten>
      ```
    - /etc/security/limits.conf:
      ```
      oracle soft memlock 41943040
      oracle hard memlock 41943040
      ```
    - Falls systemd Service benutzt wird:
      ```
      LimitMEMLOCK=infinity
      LimitNOFILE=65535+
      ```

Big 4: CPU / Memory / Disk IO / Network

- Memory: Hugepages aktivieren!
  - Alert Log:

```
 Sys-V shared memory will be used for creating SGA
 ************************************************
2019-11-07T16:59:26.560538+01:00
**********************************************************************
2019-11-07T16:59:26.562147+01:00
Dump of system resources acquired for SHARED GLOBAL AREA (SGA)
2019-11-07T16:59:26.565205+01:00
 Per process system memlock (soft) limit = 128G
2019-11-07T16:59:26.566767+01:00
 Expected per process system memlock (soft) limit to lock
 instance MAX SHARED GLOBAL AREA (SGA) into memory: 36G
2019-11-07T16:59:26.569983+01:00
 Available system pagesizes:a
  4K, 2048K
2019-11-07T16:59:26.573274+01:00
 Supported system pagesize(s):
2019-11-07T16:59:26.574826+01:00
  PAGESIZE  AVAILABLE_PAGES  EXPECTED_PAGES  ALLOCATED_PAGES  ERROR(s)
2019-11-07T16:59:26.576337+01:00
        4K       Configured               7                7        NONE
2019-11-07T16:59:26.578983+01:00
     2048K            18433           18433            18433        NONE
```

wirklich freie / unbenutzte HP =
HugePages_Free - HugePages_Rsvd

```
cat /proc/meminfo | grep -i Huge
AnonHugePages:         0 kB
ShmemHugePages:        0 kB
HugePages_Total:   18433
HugePages_Free:       56
HugePages_Rsvd:       56
HugePages_Surp:        0
Hugepagesize:       2048 kB
```

Big 4: CPU / Memory / Disk IO / Network

- Memory: Transparent Hugepages
  - Transparent (Anonymous) Hugepages haben in der Vergangenheit zu Problemen geführt:

ALERT: Disable Transparent HugePages on SLES11, RHEL6, RHEL7, OL6, OL7, and UEK2 and above (Doc ID 1557478.1)

Because Transparent HugePages are known to cause unexpected node reboots and performance problems with RAC, Oracle strongly advises to disable the use of Transparent HugePages. In addition, Transparent Hugepages may cause problems even in a single-instance database environment with unexpected performance problems or delays. As such, Oracle recommends disabling Transparent HugePages on all Database servers running Oracle.

Status:

```
cat /sys/kernel/mm/transparent_hugepage/enabled
always madvise [never]
```

RHEL 7:

```
cat /proc/cmdline
BOOT_IMAGE=/vmlinuz-3.10.0-957.el7.x86_64 root=/dev/mapper/vg00-lvol1 ro
transparent_hugepage=never ...
```

OL 7:

```
cat /proc/cmdline
BOOT_IMAGE=/vmlinuz-4.14.35-1902.6.6.el7uek.x86_64 root=/dev/mapper/ol_olymp-root
ro crashkernel=no transparent_hugepage=never
```

Big 4: CPU / Memory / Disk IO / Network

- Memory: Troubleshooting Oracle DB Startup Issue
  - **Problem:** sporadisch werden Shared Memory Segments nicht freigegeben

    Symptom: Beim Neustarten der Datenbank erhält der DBA diese Fehlermeldung:

```
ORA-32004: obsolete or deprecated parameter(s) specified for RDBMS instance
ORA-27125: unable to create shared memory segment
Linux-x86_64 Error: 28: No space left on device
Additional information: 3773
```

    Analyse: SHM Segments der Datenbank sind noch vorhanden, obwohl DB down

```
ipcs -a | grep ora
------ Shared Memory Segments --------
key        shmid      owner      perms      bytes      nattch     status
0x00000000 337117195  oracle     600        34292629504 1          dest
0x00000000 337149964  oracle     600        62914560   1          dest
0x00000000 337182733  oracle     600        40960      1          dest
```

nattch: nur 1 prozess attached

    Frage: Welcher Prozess hat diese SHM segments noch geöffnet?

```
sudo /usr/sbin/lsof  |grep 337117195
ora_j000_  1138    ora11  DEL     REG    0,12      828735530  /SYSV00000000

ps -ef |grep 1138
oracle     1138     1   0 Aug10 ?     00:00:00 ora_j000_PROD   <<<<<<<<<<<
kill -9 1138
```

Big 4: CPU / Memory / Disk IO / Network

- Disk IO:
  - Linux Disk IO Scheduler – kann I/O requests umreihen und mergen
    - Algorithmen:
      - cfq (Consistently Fair Queueing): default in RHEL7 für SATA Disks
      - deadline: empfohlen für Oracle ASM*
      - noop
    - Prüfung:
      - cat /sys/block/sda/queue/scheduler
        noop [deadline] cfq
  - init.ora:
    - filesystemio_options = SETALL
    - disk_asynch_io = TRUE

direct IO:
Umgehen des Filesystem Caches / Vermeidet doppeltes Caching

async IO:
Prozesse können mehrere I/O Requests absetzen und in der Zwischenzeit mit anderen Instruktionen fortsetzen

* : https://docs.oracle.com/en/database/oracle/oracle-database/19/cwlin/setting-the-disk-io-scheduler-on-linux.html#GUID-B59FCEFB-20F9-4E64-8155-7A61B38D8CDF)

Big 4: CPU / Memory / Disk IO / Network

Network

- /etc/sysconfig/network-scripts/ifcfg-<interface>
- Zeitsynchronisierung: chronyd statt ntpd
- DNS: /etc/resolv.conf
- nslookup <servername>
- Network-Interfaces: /sbin/ifconfig –a
- Routing Table: netstat –nr
- Bonding: mehrere physische Interfaces werden zu einem virtuellen Interface zusammengefasst (active/active oder active/passive):
  cat /proc/net/bonding/bond0
- 10 GbE / 25 GbE / 40 GbE / 100 GbE
- Früher öfters Probleme bei Autonegotiation von Speed und Duplex (Full/Half)
- ethtool <interface>

Big 4: CPU / Memory / Disk IO / Network

- Network
  - Früher öfters Probleme bei Autonegotiation von Speed und Duplex (Full/Half)
  - ethtool <interface>

```
# ethtool enp3s0
Settings for enp3s0:
                Supported ports: [ TP ]
                Supported link modes:   10baseT/Half 10baseT/Full
                                        100baseT/Half 100baseT/Full
                                        1000baseT/Full
                Supported pause frame use: Symmetric
                Supports auto-negotiation: Yes
                Supported FEC modes: Not reported
                Advertised link modes:  10baseT/Half 10baseT/Full
                                        100baseT/Half 100baseT/Full
                                        1000baseT/Full
                Advertised pause frame use: Symmetric
                Advertised auto-negotiation: Yes
                Advertised FEC modes: Not reported
                Speed: 1000Mb/s
                Duplex: Full
                Port: Twisted Pair
                PHYAD: 1
                Transceiver: internal
                Auto-negotiation: on
                MDI-X: off (auto)
                Supports Wake-on: pumbg
                Wake-on: g
                Current message level: 0x00000007 (7)
                                                   drv probe link
                Link detected: yes
```

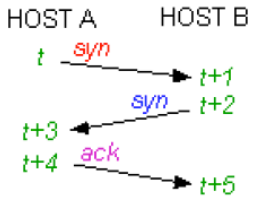
Big 4: CPU / Memory / Disk IO / Network

Network

- Anzeichen für blockierende Firewall: Socket bleibt im SYN_SENT status

```
netstat -an | grep SYN
tcp        0      1 10.1.1.1:45904          8.8.8.8:80          SYN_SENT
```

- Three-Way Handshake

Time	Event	DIAGRAM
t	**Host A sends** a TCP **SYN**chronize packet to Host B	
t+1	Host B receives A's **SYN**	
t+2	**Host B sends** it's own **SYN**chronize	
t+3	Host A receives B's **SYN**	
t+4	**Host A sends ACK**nowledge	
t+5	Host B receives **ACK**. ***TCP connection is established.***	

Big 4: CPU / Memory / Disk IO / Network

- Network: Network Sniffing mittels tcpdump / wireshark

```
SQL> alter user scott identified by tiger;
SQL> select name, password, spare4 from sys.user$ where name = 'SCOTT'
NAME       PASSWORD   SPARE4
---------- ---------- ----------------------------------------------------------------------------
SCOTT                 S:DC90F997ADAA46B9F1C228F4BE3E97412969EFC6C29CE26FE3E7EE2942A8;T:F4A796EC26...
```

#tcpdump -A -nn -tt -i <interface> port 1521 and host 192.168.0.61

```
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on vboxnet1, link-type EN10MB (Ethernet), capture size 262144 bytes
1573067571.870831 IP 192.168.0.1.37824 > 192.168.0.61.1521: Flags [P.], seq 2094271333:2094271730, ack 3313303054, win 302, options [nop,nop,TS val 2824454858 ecr 2979925976], length 397
E...r.@.@.D........=....|..e.|....../2.....
.Y.................i    .....................^
!..............l...........................................................5...........................................
..........................................................$alter user scott identified by tiger...........................................
1573067571.899567 IP 192.168.0.61.1521 > 192.168.0.1.37824: Flags [P.], seq 1:190, ack 397, win 323, options [nop,nop,TS val 2979991607 ecr 2824454858], length 189
E...d@@.@.T8...=.........|..|......CE......
...7.Y................c.................................................+.........................
......6..................[<..........................................................
1573067571.899592 IP 192.168.0.1.37824 > 192.168.0.61.1521: Flags [.], ack 190, win 325, options [nop,nop,TS val 2824454887 ecr 2979991607], length 0
E..4r.@.@.Fm.......=....|....|.....E.......
.Y.....7
1573067591.199542 IP 192.168.0.1.37824 > 192.168.0.61.1521: Flags [P.], seq 397:823, ack 190, win 325, options [nop,nop,TS val 2824474188 ecr 2979991607],
length 426
E...r.@.@.D........=....|....|.....E.......
.Z.L...7..........i......................^.a.........................................................5.......................
.......................................................................................Aselect name, password, spare4 from sys.user$ where
name = 'SCOTT'.................................................
1573067591.200936 IP 192.168.0.61.1521 > 192.168.0.1.37824: Flags [P.], seq 190:960, ack 823, win 342, options [nop,nop,TS val 2980010908 ecr
2824474188], length 770
E..6dA@.@.Q....=.........|..|......V.......
..[..Z.L...............R.(=.
..=R@Tc.[.xw.............\.....................................i........?.........NAME...........................................i........?.........PASSWORD.............................
...i........?.........SPARE4..................xw.........................."...............................SCOTT..S:DC90F
997ADAA46B9F1C228F4BE3E97412969EFC6C29CE26FE3E7EE2942A8;T:F4A7967AFC06C5C803125890F79AEC265841AA59EE6A0A022EBA838CBE3E
2F3EEC383B2D56E3D4B391E6DC932D4A53BB901CDD7BCA13C2616121DBD9829C158AF0A89FC3B7D6A21CED031D152C469DF0.....c.........................
.......
```

Troubleshooting mit "strace"

- Grid Infrastructure Upgrade von 12.1.0.2 auf 19.5.0 bricht bei "rootupgrade.sh" mit Fehler ab

```
2019/11/07 11:21:49 CLSRSC-117: Failed to start Oracle Clusterware stack
Died at /opt/oracle/gridl1/19.0.0.0/grid/crs/install/crsupgrade.pm line 1586.
```

- SEV1 Service Request
- Support / Development bei der Fehlersuche, ein paar non-public Bugs werden referenziert
- GI benutzt HAIP IP´s. Bei Upgrade auf 19c ändert sich "Subnet Mask" dieser IP.
- Da rollierendes Upgrade funktionieren soll, muß IP von neuem Subnet (19c) mit IP von altem Subnet (12.1) kommunizieren können
- Dafür ist eine Netzwerk-Route nötig. Diese wird im Zuge des rootupgrade.sh von ohasd_orarootagent_root.trc beim Startup von der HAIP Resource gesetzt.

ohasd_orarootagent_root.trc:

```
2019-11-07 11:10:50.753 : USRTHRD:3739182848: [ INFO] {0:5:3} Thread:[NetHAMain] HAIP: add routedata 169.254/16 / 10.19.240.40 / 255.255.0.0 / eth1.882 / eth1.882
2019-11-07 11:10:50.759 : USRTHRD:3739182848: [ INFO] {0:5:3} Thread:[NetHAMain] Add HAIP route failed 169.254/16 / 255.255.0.0 / 10.19.240.40 / eth1.882
2019-11-07 11:10:50.759 : USRTHRD:3739182848: [ INFO] {0:5:3} (null) category: -2, operation: routeadd, loc: system, OS error: 0, other: failed to execute route add, ret 256
```

failed

- Tracing mit "strace"

```
PID=$(cat /opt/oracle/gridbase/crsdata/node1/output/ohasd_orarootagent_root.pid)
strace -tt -f -o /opt/oracle/gridbase/strace_routadd.txt -s 100  -p $PID
```

```
14545 11:10:50.754217 execve("/bin/sh", ["sh", "-c", "/sbin/ip route add 169.254/16/16 dev eth1.882"], [/* 67 vars */] <unfinished ...>
14545 11:10:50.758248 execve("/sbin/ip", ["/sbin/ip", "route", "add", "169.254/16/16", "dev", "eth1.882"], [/* 68 vars */]) = 0
14545 11:10:50.759711 write(2, "Error: any valid prefix is expected rather than \"169.254/16/16\".\n", 65) = 65
```

Richtige Syntax: `/sbin/ip route add 169.254.0.0/16 dev eth1.882`

BUG 30523200 - ASM1 FAILS WHEN UPGRADE 12.1 TO 19C FAILS WITH ONE HAIP

Demos

- Demo 1: Exploring Linux Host (ca. 8 min)
- Demo 2: Exploring Oracle Environment (ca. 7 min)
- Demo 3: Performance Monitoring Tools (ca. 5 min)
- Demo 4: strace – tnsping (ca. 1 min)
- Demo 5: strace – startup db (ca. 3 min)
- Demo 6: sysdig (ca. 3 min)
- Demo 7: execsnoop (ca. 1 min)
- Demo 8: perf profiling - flamegraph (ca. 5 min)

Link-Sammlung

- https://de.wikipedia.org/wiki/Liste_von_Linux-Distributionen
- https://de.wikipedia.org/wiki/Geschichte_von_Linux
- https://de.wikipedia.org/wiki/Red_Hat_Enterprise_Linux
- https://blogs.oracle.com/scoter/oracle-linux-and-unbreakable-enterprise-kernel-uek-releases
- https://access.redhat.com/documentation/en-us/reference_architectures/2017/html-single/deploying_oracle_database_12c_release_2_on_red_hat_enterprise_linux_7/index
- https://utcc.utoronto.ca/~cks/space/blog/linux/KernelMemoryZones
- https://linux-mm.org/Low_On_Memory
- http://ptgmedia.pearsoncmg.com/images/0131453483/downloads/gorman_book.pdf
- https://eklitzke.org/swappiness
- https://fritshoogland.wordpress.com/2017/07/25/linux-memory-usage
- https://medium.com/@FranckPachot/proc-meminfo-formatted-for-humans-350c6bebc380
- https://savvinov.com/2019/10/14/memory-fragmentation-the-silent-performance-killer/
- https://github.com/brendangregg/perf-tools/blob/master/execsnoop
- http://www.brendangregg.com/blog/2017-08-08/linux-load-averages.html
- http://www.brendangregg.com/perf.html#TimedProfiling
- https://blog.tanelpoder.com/psnapper/
- https://sysdig.com/
- https://blog.tanelpoder.com/seminar/practical-linux-performance-application-troubleshooting-training/

Q & A

Martin Decker
ora-solutions.net

E-Mail: martin.decker@ora-solutions.net
Internet: http://www.ora-solutions.net
Blog: http://www.ora-solutions.net/web/blog/