# Schweben auf Wolke 7 - Oracle Linux 7.x unter der Lupe

**Ralf Germann**
**Trivadis AG**
**Glattbrugg (Schweiz)**

**Schlüsselworte**
Oracle Linux 7.0, RedHat Enterprise Linux 7.0, OL 7.0, RHEL 7.0, Anaconda Installer, Kickstart, Text Installer, XFS, systemd, Ablösung sysV, GRUB2, LVM Cache, pNFS, BTRFS, kpatch, ksplice, Swap Memory Compression, USB 3.0, KVM, Linux Containers, Pacemaker, keepalived, HAProxy, Network-Teaming, firewalld, Ablösung iptables, 40GBit, Samba 4.1.0, UEK R3, Naming Network-Devices, pykickstart, nmcli, Control Groups

**Einleitung**
Unter Systemadministratoren bereits im Vorfeld heiß diskutiert, habe auch ich mich sehnsüchtig auf Oracle Linux 7 gefreut. Kaum waren die ersten testbaren Versionen verfügbar, wurden diese auch sogleich installiert und ausprobiert. Bald war jedoch klar, dass der Unterschied zwischen Version 6.x und 7.x gigantisch ist und keinesfalls mit der Ablösung von 5.x durch 6.x verglichen werden kann. Selbst erfahrene Administratoren müssen wohl oder übel zurück ans „Zeichenbrett“ und sich mit mehr Zeit der Materie widmen, als sie es sich möglicherweise von anderen Versionswechseln gewohnt waren.

Es fängt mit der Installation an, egal ob grafisch, textbasiert oder mittels Kickstart automatisiert. XFS, BTRFS sowie die neuen Möglichkeiten des LVM (z.B. der Cache) leisten genauso ihren Beitrag zum Mehraufwand wie die Ablösung von sysV durch systemd. Einfache Änderungen am Netzwerk, des Hostnamen oder der GRUB-Konfiguration sind nicht mehr auf die bekannte Art und Weise möglich. Dann folgen Änderungen, die ich auf den ersten Blick hinterfragt habe, wie z.B. die Swap Memory Compression oder die Alternative für Bonding, genannt Network-Teaming. Und zu guter Letzt hat mich auch noch der neue UEK-Kernel beim Updaten geärgert.

Im ersten Moment fühlte ich mich darum ziemlich überfahren. Nach einer kleineren Pause habe ich mit einer etwas offeneren Einstellung nochmals den Neuerungen gewidmet und durfte erfreulicherweise meine Meinung in einigen Punkten revidieren.

**Die Installation – Strenger, einfacher und übersichtlicher?!**

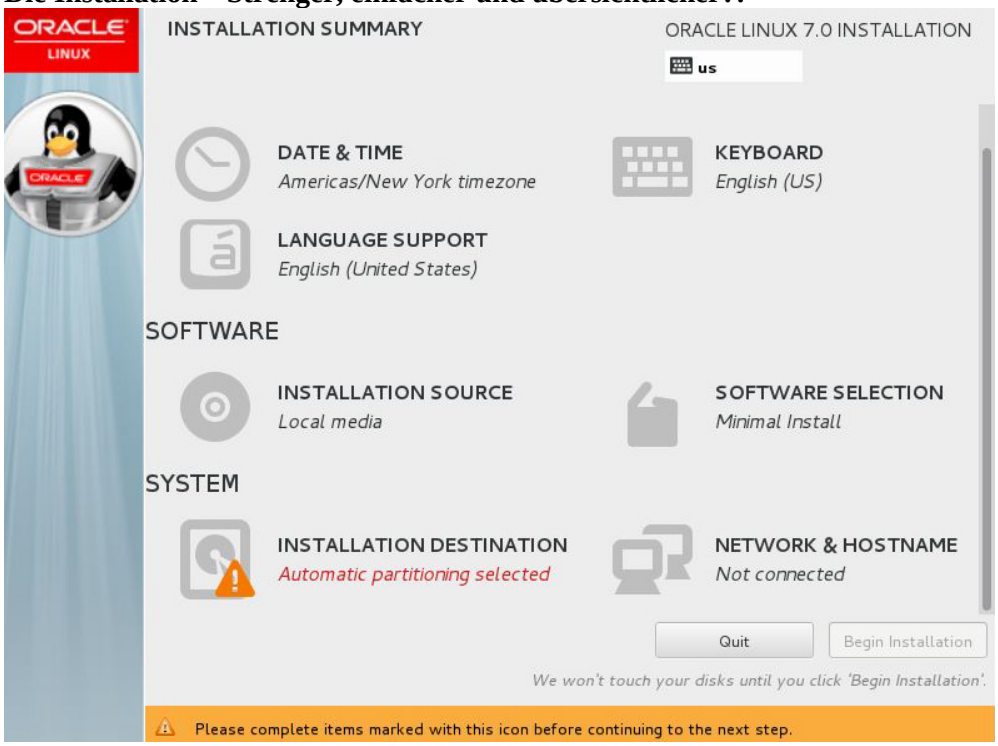


*Abb. 1: Der Anaconda Installer GUI*

Der Anaconda Installer – sowohl grafisch als auch textbasiert – wurde komplett überarbeitet. Der Text Installer wurde sogar von Grund auf neu geschrieben. Den etwas verstaubten Wizard hat man durch ein Fenster bzw. durch eine Seite ersetzt von wo aus praktisch alle Einstellungen zentral vorgenommen werden können. Ausnahme bilden hier die Sprachauswahl für die Installation sowie das Setzen des root Passworts und die Erstellung zusätzlicher Benutzer.

Dies mag zugegebenermaßen grafisch tatsächlich übersichtlicher und moderner wirken, jedoch sehne ich mir beim nun sehr unübersichtlichen Text Installer den Wizard zurück.

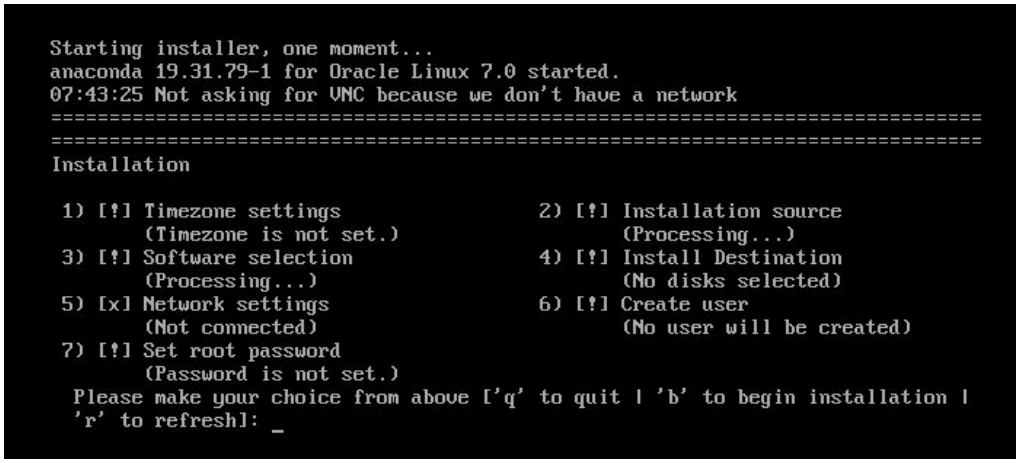


*Abb. 2: Der Text Installer*

Besitzt der Server eine öffentliche IP-Adresse werden gewisse Einstellungen automatisch mit einem Vorschlag versehen. Hierzu gehören z.B. die Keyboard- und Spracheinstellungen als auch die Zeitzone. Bei internen IP-Adressen funktionieren diese Vorschläge natürlich nicht und die Einstellungen müssen manuell vorgenommen werden. **ACHTUNG**, hier lauert bereits die erste Stolperfalle. Die NTP-Funktion, welche sich im ersten Optionsfeld befindet kann erst eingeschaltet werden, nachdem das Netzwerk-Interface im letzten (!) Optionsfeld eingeschaltet wurde. Ansonsten bleibt der Schieberegler unbeeindruckt grau und gibt auch keinen Fehler aus.

Ansonsten sind die anderen Optionsfelder mehr oder weniger wie aus früheren Installationen gewohnt zu benutzen. Per Default wird allerdings nicht mehr ext4, sondern xfs als Filesystem verwendet und in den LVM-Einstellungen kann das in Version 6.4 eingeführte „thin provisioning" für LVM verwendet werden. Bei der minimalen Installation gilt es zu beachten, dass kein perl und auch keine net-tools (ifconfig) installiert werden. Natürlich können diese bei Bedarf nachinstalliert werden.

Praktisch identisch mit 6.x ist die Installation mit einem automatisierten Kickstart-File. Hier wurden augenscheinlich nur wenige Änderungen vorgenommen. Direkt aufgefallen ist die Änderung von utc auf isutc in den timezone-Einstellungen und die sehr strenge Handhabung mit dem `%end` Tag bei den Sektionen packages, pre und post. Bei früheren Versionen wurde das Vergessen dieses Abschluss-Tags meist ignoriert und die Installation trotzdem gestartet. Bei der Version 7.x wird dies allerdings streng geprüft und die Installation bei einem Fehlen des Tags abgebrochen. Mit dem Tool `pykickstart(1M)` hat man allerdings im Vorfeld die Möglichkeit ein Kickstart-File auf seine Richtigkeit zu prüfen. Ebenfalls mit dabei ist die Möglichkeit Änderungen zwischen einzelnen Kickstartversionen anzeigen zu lassen. **ACHTUNG:** Syntax-Änderungen sind nicht aufgeführt.

Prüfung des Kickstartfiles:
```
ksvalidator /home/trivadis/OL7/kickstart.cfg
```

Aufzeigen der Unterschiede zwischen RHEL6 und RHEL7:
```
ksverdiff --from RHEL6 --to RHEL7
```

Gefühlt dauert die Installation (v.a. die Post-Installation) etwas länger als früher, dies fällt aber kaum ins Gewicht und ist nach der merklich kürzeren Aufstartzeit des Betriebssystems auch gleich wieder vergessen.

**XFS – Nicht nur etwas für Datenkraken**
Das von Silicon Graphics entwickelte Filesystem XFS hat nun auch seinen Einzug in Oracle Linux gehalten und wurde sogleich zum Standardfilesystem erklärt. Es handelt sich hierbei um ein Journaling Filesystem, welches theoretisch eine Größe von bis zu 16EB (16 Millionen TB) haben kann. Die maximale Größe pro Datei ist auf 8EB limitiert genauso wie die Directory-Struktur mit maximal 10 Millionen Einträgen. Die Limitierungen wurden aber innerhalb von OL7 nochmals etwas angepasst: Max. Volumegröße: 500TB und Max. Größe pro File: 16TB.

| Limit | EXT3 | EXT4 | XFS |
|---|---|---|---|
| Max. FS Grösse | 16TB | 16TB | 16EB |
| Max Filegrösse | 2TB | 16TB | 8EB |
| Max. Extentgrösse | 4kB | 128MB | 8GB |
| Max. Inodes | $2^{32}$ | $2^{32}$ | $2^{64}$ |

Zum Thema Performance gibt es zahlreiche Benchmark-Tests mit unterschiedlichen Ergebnissen. Zusammengefasst kann man sagen, dass XFS vor allem dann punktet, wenn es um sehr große Datenumgebungen geht.

Ein Vorteil von XFS ist aber sicherlich die Möglichkeit das Filesystem online zu defragmentieren und auch zu vergrößern. Leider ist eine Verkleinerung zum jetzigen Zeitpunkt nicht möglich. Meiner Meinung nach ist auch bei der Vergrößerung der Umstand, dass diese durch Angabe von Größe in Blöcken gemacht werden muss, nicht unbedingt ideal.

Ebenfalls Punkte sammeln kann XFS mit der Möglichkeit ein Filesystem parallel auf mehrere andere Filesysteme zu kopieren und dumps auf andere Speicherorte machen zu können. Bei beiden Aktionen müssen jedoch die betroffenen Filesysteme im ungemounteten Zustand sein.

Beispiel Backup von /test nach /copy1, /copy2 und /copy3:

```
umount /test
umount /copy1
umount /copy2
umount /copy3
xfs_copy /dev/mapper/vg_root-lv_test /dev/mapper/vg_root-lv_copy1
/dev/mapper/vg_root-lv_copy2 /dev/mapper/vg_root-lv_copy3
```

Beispiel FS-dump von /copy3 nach /test/metadump_copy3 erstellen und wiedereinstpielen:

```
umount /copy3
xfs_metadump /dev/mapper/vg_root-lv_copy3 /test/metadump_copy3
mount /copy3
rm –rf /copy3/*
umount /copy3
xfs_mdrestore /test/metadump_copy3 /dev/mapper/vg_root-lv_copy3
mount /copy3
```

**LVM-Cache – Der Booster für Ihre Filesysteme**

Neu gibt es die Möglichkeit innerhalb einer VG einen sogenannten LVM-Cache hinzuzufügen. Dies macht v.a. bei sehr großen VGs (SAN oder DAS) Sinn, sofern es sich beim Cache-Device um ein sehr schnelles Device (z.B. SSD, FlashStorage, RAID 0) handelt. Um den Cache zu erstellen wird eine LV für die Metadaten des Caches und eine LV für den eigentlichen Cache benötigt, welche idealerweise auf unterschiedliche physische Devices erstellt werden sollten. Beide LVs werden zu einem Pool zusammengefasst welcher dann wiederum einer LV mit Filesystemstruktur zugewiesen wird.

**ACHTUNG:** Folgende Einschränkungen sind in Bezug auf den LVM-Cache zu beachten:

- Die Cache-LVs können vergrössert oder verkleinert werden.
- Bei einem `pvmove` wird der Cache-Pool nicht verschoben.
- Ein `vgsplit` ist nicht möglich, wenn eine VG einen oder mehrere LVM-Cache(s) enthält.

Beispiel Erstellung eines Cache-Pools für lv_test:

```
lvcreate -L 1G -n lv_cache_meta vg_root /dev/sdb
lvcreate -L 3.5G -n lv_cache vg_root /dev/sdc
lvconvert --type cache-pool --poolmetadata vg_root/lv_cache_meta
vg_root/lv_cache
lvconvert --type cache --cachepool vg_root/lv_cache vg_root/lv_test
```

**Ablösung sysV durch systemd**
Eine der größten Neuerungen ist sicherlich die Einführung von systemd. Dieses löst sysV ab und damit auch `init(1M)`. systemd ist der erste Prozess beim Start und der letzte Prozess beim Shutdown. Da mit systemd Prozesse gleichzeitig geladen werden können, dürfte diese Änderung der Hauptgrund für den schnelleren Bootvorgang sein. Neu werden mit `systemctl(1M)` Befehle wie Service stop oder start ausgeführt:

| Aktion | Systemctl-Befehl | „Alter“ Befehl |
|---|---|---|
| Services anzeigen | systemctl list-unit-files | chkconfig –list |
| Service (de)aktivieren | systemctl disable / enable xyz | chkconfig xyz off / on |
| Shutdown | systemctl poweroff | poweroff |
| Neustart | systemctl reboot | reboot |
| Starten / Stoppen / Status von Services | systemctl start /stop / status xyz | service xyz start / stop /status |

**Und was gibt es sonst noch?**
Leider würde es den Rahmen sprengen sämtliche neuen Features detailliert in diesem Manuskript niederzuschreiben, weshalb andere nennenswerte Features nur im Schnelldurchlauf behandelt werden.

Mit kpatch hat RedHat nachgezogen und eine Alternative zu ksplice von Oracle entwickelt, welche es ermöglicht einen Kernel online zu patchen.

USB 3.0 wird nun von KVM unterstützt, genauso wie Windows 8, 8.1 und Server 2012 (R2) als Gastbetriebssystem. Ausserdem können VPC und VHDX Disks readonly angebunden werden.

Pacemaker ist neu der Default für das Clustering. Der Loadbalancer Piranha wurde duch keepalived und HAProxy (TCP/http resource proxy) abgelöst.

Mit network-teaming wird eine Alternative für Bonding eingeführt. Mit dem Tool bond2team lässt sich die Migration leichter bewältigen.

Iptables erhält ein neues „GUI“ namens firewalld:
`firewall-cmd [OPTION(S)] SUBCOMMAND`

Samba 4.1 und der Support von pNFS haben Einzug in OL7 gehalten.

GRUB2 ist nun auch unter OL7 der Standard. PowerPC wird supported, genauso wie HFS+ von Apple und NTFS. Wird der UEK R3-Kernel eingesetzt, ist ein UEFI Secure Boot nicht möglich.

**Wie migriere ich?**
Meine Empfehlung: Durch eine Neuinstallation. Dies erspart Ihnen ziemlich sicher viel Ärger. Unter folgenden Bedingungen wäre es allerdings möglich ein Upgrade von 6.x auf 7.x durchzuführen:

- Die minimalen Installationsanforderungen sind erfüllt
- UEK R3 war bereits auf dem 6.x System installiert und der default Kernel (Upgrade von UEK R2 wird nicht unterstützt)
- Kein Oracle Produkt ist auf dem 6.x System installiert
- Der Upgrade wird nur für Systeme mit der minimalen Installationsumgebung unterstützt.

**Kontaktadresse:**
Ralf Germann
Consultant
Trivadis AG
Europa-Straße 5
CH-8152 Glattbrugg (Zürich)

| | |
|---|---|
| Telefon: | +41 (0) 76-337 1983 |
| Fax: | +41 (0) 44-808 7021 |
| E-Mail | ralf.germann@trivadis.com |
| Internet: | www.trivadis.com |