

>>> BACKUPS UNTER LINUX

>>> ...denn Murphy kommt bestimmt...

Name: Florian Knodt / adlerweb[†] (LUG Mayen-Koblenz)

Datum: 2016-04-14

[†]lugmyk@adlerweb.info · www.adlerweb.info · www.lug-myk.org

>>> Anforderungen

* Mehrere Quellsysteme

>>> Anforderungen

- * Mehrere Quellsysteme
- * Ein Zielsystem

>>> Anforderungen

- * Mehrere Quellsysteme
- * Ein Zielsystem
- * Teilweise sehr langsames/instabiles/unsicheres Netzwerk dazwischen

>>> Anforderungen

- * Mehrere Quellsysteme
- * Ein Zielsystem
- * Teilweise sehr langsames/instabiles/unsicheres Netzwerk dazwischen
- * Viele Duplikate

>>> Anforderungen

- * Mehrere Quellsysteme
- * Ein Zielsystem
- * Teilweise sehr langsames/instabiles/unsicheres Netzwerk dazwischen
- * Viele Duplikate
- * Schneller Zugriff auf einzelne Dateien in der Sicherung

>>> Anforderungen

- * Mehrere Quellsysteme
- * Ein Zielsystem
- * Teilweise sehr langsames/instabiles/unsicheres Netzwerk dazwischen
- * Viele Duplikate
- * Schneller Zugriff auf einzelne Dateien in der Sicherung
- * Revisionen der letzten x Tage beibehalten


```
>>> rdiff-backup1
```

¹<http://www.nongnu.org/rdiff-backup/>

```
>>> rdiff-backup1
```

+ Rdiff/rsync: Überträgt und speichert nur Änderungen

¹<http://www.nongnu.org/rdiff-backup/>

```
>>> rdiff-backup1
```

- + Rdiff/rsync: Überträgt und speichert nur Änderungen
- + Läuft über SSH

¹<http://www.nongnu.org/rdiff-backup/>

>>> `rdiff-backup`¹

- + Rdiff/rsync: Überträgt und speichert nur Änderungen
- + Läuft über SSH
- + 1:1-Kopie direkt am Backupziel zugreifbar

¹<http://www.nongnu.org/rdiff-backup/>

>>> `rdiff-backup`¹

- + Rdiff/rsync: Überträgt und speichert nur Änderungen
- + Läuft über SSH
- + 1:1-Kopie direkt am Backupziel zugreifbar
- + Ältere Revisionen von Dateien über Befehl abrufbar

¹<http://www.nongnu.org/rdiff-backup/>

>>> `rdiff-backup`¹

- + Rdiff/rsync: Überträgt und speichert nur Änderungen
- + Läuft über SSH
- + 1:1-Kopie direkt am Backupziel zugreifbar
- + Ältere Revisionen von Dateien über Befehl abrufbar
- Python 2

¹<http://www.nongnu.org/rdiff-backup/>

>>> `rdiff-backup`¹

- + Rdiff/rsync: Überträgt und speichert nur Änderungen
- + Läuft über SSH
- + 1:1-Kopie direkt am Backupziel zugreifbar
- + Ältere Revisionen von Dateien über Befehl abrufbar
- Python 2
- Letztes update 2009

¹<http://www.nongnu.org/rdiff-backup/>

>>> `rdiff-backup`¹

- + Rdiff/rsync: Überträgt und speichert nur Änderungen
- + Läuft über SSH
- + 1:1-Kopie direkt am Backupziel zugreifbar
- + Ältere Revisionen von Dateien über Befehl abrufbar
- Python 2
- Letztes update 2009
- Keine Kompression oder Deduplizierung

¹<http://www.nongnu.org/rdiff-backup/>

>>> BitTorrent Sync², Owncloud³, Sparkleshare⁴

²<https://www.getsync.com/> (Closed source)

³<https://owncloud.org/>

⁴<http://sparkleshare.org/>

>>> BitTorrent Sync², Owncloud³, Sparkleshare⁴

Desktop-Clients zur Synchronisation mit Server oder anderen Clients

²<https://www.getsync.com/> (Closed source)

³<https://owncloud.org/>

⁴<http://sparkleshare.org/>

>>> BitTorrent Sync², Owncloud³, Sparkleshare⁴

Desktop-Clients zur Synchronisation mit Server oder anderen Clients

– Eher für Desktopbetrieb und reine Replikation

²<https://www.getsync.com/> (Closed source)

³<https://owncloud.org/>

⁴<http://sparkleshare.org/>

>>> BitTorrent Sync², Owncloud³, Sparkleshare⁴

Desktop-Clients zur Synchronisation mit Server oder anderen Clients

- Eher für Desktopbetrieb und reine Replikation
- Probleme mit großen Datenmengen

²<https://www.getsync.com/> (Closed source)

³<https://owncloud.org/>

⁴<http://sparkleshare.org/>

>>> BitTorrent Sync², Owncloud³, Sparkleshare⁴

Desktop-Clients zur Synchronisation mit Server oder anderen Clients

- Eher für Desktopbetrieb und reine Replikation
- Probleme mit großen Datenmengen
- Augenmerk auf Daten, weniger die Metadaten (ACL, Attribute, etc)

²<https://www.getsync.com/> (Closed source)

³<https://owncloud.org/>

⁴<http://sparkleshare.org/>

```
>>> Unison5
```

⁵<http://www.cis.upenn.edu/~bcpierce/unison/>

>>> Unison⁵

Open source user-level file-synchronization tool für OSX, Unix, und Windows mit Zweiwegereplikation und inkrementeller Übertragung.

⁵<http://www.cis.upenn.edu/~bcpierce/unison/>

>>> Unison⁵

Open source user-level file-synchronization tool für OSX, Unix, und Windows mit Zweiwegereplikation und inkrementeller Übertragung.

- Revisionsverwaltung nur schwer aufräumbar

⁵<http://www.cis.upenn.edu/~bcpierce/unison/>

>>> Unison⁵

Open source user-level file-synchronization tool für OSX, Unix, und Windows mit Zweiwegereplikation und inkrementeller Übertragung.

- Revisionsverwaltung nur schwer aufräumbar
- OCalm?!

⁵<http://www.cis.upenn.edu/~bcpierce/unison/>

>>> Unison⁵

Open source user-level file-synchronization tool für OSX, Unix, und Windows mit Zweiwegereplikation und inkrementeller Übertragung.

- Revisionsverwaltung nur schwer aufräumbar
- OCalm?!
- Aus OSX-Ecke? -> Probleme mit Metadaten

⁵<http://www.cis.upenn.edu/~bcpierce/unison/>

>>> Obnam⁹

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

+ Komprimiert (Deflate⁶)

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

+ Komprimiert (Deflate⁶)

+ Dedupliziert

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

+ Komprimiert (Deflate⁶)

+ Dedupliziert

+ Verschlüsselt (GnuPG)

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

- + Komprimiert (Deflate⁶)
- + Dedupliziert
- + Verschlüsselt (GnuPG)
- + SFTP

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

+ Komprimiert (Deflate⁶)

+ Dedupliziert

+ Verschlüsselt (GnuPG)

+ SFTP

o Wird aktiv entwickelt⁷

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

+ Komprimiert (Deflate⁶)

+ Dedupliziert

+ Verschlüsselt (GnuPG)

+ SFTP

o Wird aktiv entwickelt⁷

- Python 2⁸

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

- + Komprimiert (Deflate⁶)

- + Dedupliziert

- + Verschlüsselt (GnuPG)

- + SFTP

- o Wird aktiv entwickelt⁷

- Python 2⁸

- Bei mir recht langsam via SFTP

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Obnam⁹

+ Komprimiert (Deflate⁶)

+ Dedupliziert

+ Verschlüsselt (GnuPG)

+ SFTP

o Wird aktiv entwickelt⁷

- Python 2⁸

- Bei mir recht langsam via SFTP

- Hier teilweise Fehler nach Netzwerkabbrüchen während Backups

⁶aka ZIP

⁷sort of...

⁸mal wieder..

⁹<http://obnam.org/>

>>> Was ist BORG¹⁰

Ebenfalls eine Software speziell für Datensicherungen. Schlanke Systemvoraussetzungen, professionelle Funktionen und aktive Weiterentwicklung.

¹⁰<http://borgbackup.readthedocs.org/en/stable/>

>>> Was ist BORG¹⁰

Ebenfalls eine Software speziell für Datensicherungen. Schlanke Systemvoraussetzungen, professionelle Funktionen und aktive Weiterentwicklung.

+ Interne Revisionsverwaltung um auf ältere Datenstände zugreifen zu können

¹⁰<http://borgbackup.readthedocs.org/en/stable/>

>>> Was ist BORG¹⁰

Ebenfalls eine Software speziell für Datensicherungen. Schlanke Systemvoraussetzungen, professionelle Funktionen und aktive Weiterentwicklung.

- + Interne Revisionsverwaltung um auf ältere Datenstände zugreifen zu können
- + Ziel kann mittels FUSE gemountet werden um direkt auf die gesicherten Dateien zuzugreifen

¹⁰<http://borgbackup.readthedocs.org/en/stable/>

>>> Was ist BORG¹⁰

Ebenfalls eine Software speziell für Datensicherungen. Schlanke Systemvoraussetzungen, professionelle Funktionen und aktive Weiterentwicklung.

- + Interne Revisionsverwaltung um auf ältere Datenstände zugreifen zu können
- + Ziel kann mittels FUSE gemountet werden um direkt auf die gesicherten Dateien zuzugreifen
- + Kann nicht nur Ordner/Dateien sondern auch Partitionen oder auch ganze Laufwerke sichern

¹⁰<http://borgbackup.readthedocs.org/en/stable/>

>>> Was ist BORG¹⁰

Ebenfalls eine Software speziell für Datensicherungen. Schlanke Systemvoraussetzungen, professionelle Funktionen und aktive Weiterentwicklung.

- + Interne Revisionsverwaltung um auf ältere Datenstände zugreifen zu können
- + Ziel kann mittels FUSE gemountet werden um direkt auf die gesicherten Dateien zuzugreifen
- + Kann nicht nur Ordner/Dateien sondern auch Partitionen oder auch ganze Laufwerke sichern
- + Sicherungsziel lokal, entferntes System mit BORG (via SSH) oder auch via SSHFS/CIFS/NFS/...

¹⁰<http://borgbackup.readthedocs.org/en/stable/>

>>> Was ist BORG¹⁰

Ebenfalls eine Software speziell für Datensicherungen. Schlanke Systemvoraussetzungen, professionelle Funktionen und aktive Weiterentwicklung.

- + Interne Revisionsverwaltung um auf ältere Datenstände zugreifen zu können
- + Ziel kann mittels FUSE gemountet werden um direkt auf die gesicherten Dateien zuzugreifen
- + Kann nicht nur Ordner/Dateien sondern auch Partitionen oder auch ganze Laufwerke sichern
- + Sicherungsziel lokal, entferntes System mit BORG (via SSH) oder auch via SSHFS/CIFS/NFS/...
- + Sicherungsziel kann für mehrere Quellen verwendet werden

¹⁰<http://borgbackup.readthedocs.org/en/stable/>

>>> Geschichte

* 2015 von "Attic Backup" (2010) geforked

>>> Geschichte

- * 2015 von "Attic Backup" (2010) geforked
- * Entwicklung von Attic war den Contributern zu zäh

>>> Geschichte

- * 2015 von "Attic Backup" (2010) geforked
- * Entwicklung von Attic war den Contributern zu zäh
- * Diverse Bugs behoben und Funktionen nachgerüstet

>>> Geschichte

- * 2015 von "Attic Backup" (2010) geforked
- * Entwicklung von Attic war den Contributern zu zäh
- * Diverse Bugs behoben und Funktionen nachgerüstet
- * Nicht mehr zu Attic kompatibel

>>> Aufbau & Code

* >90% Python

>>> Aufbau & Code

* >90% Python3

>>> Aufbau & Code

- * >90% Python3

- * Performancekritische Teile in C/Cython

>>> Aufbau & Code

- * >90% Python3
- * Performancekritische Teile in C/Cython
- * Unterstützt POSIX-xattr und ACLs

>>> Aufbau & Code

- * >90% Python3
- * Performancekritische Teile in C/Cython
- * Unterstützt POSIX-xattr und ACLs
- * BSD-Lizenz

>>> Aufbau & Code

- * >90% Python3
- * Performancekritische Teile in C/Cython
- * Unterstützt POSIX-xattr und ACLs
- * BSD-Lizenz
- * Continuous Integration, Unit Tests

>>> Sicherheit

- * Alle Daten und Metadaten werden signiert - versehentliche und absichtliche Veränderung kann erkannt werden

>>> Sicherheit

- * Alle Daten und Metadaten werden signiert - versehentliche und absichtliche Veränderung kann erkannt werden
- * Daten und Metadaten werden verschlüsselt

>>> Sicherheit

- * Alle Daten und Metadaten werden signiert - versehentliche und absichtliche Veränderung kann erkannt werden
- * Daten und Metadaten werden verschlüsselt
 - > AES256-basierend

>>> Sicherheit

- * Alle Daten und Metadaten werden signiert - versehentliche und absichtliche Veränderung kann erkannt werden
- * Daten und Metadaten werden verschlüsselt
 - > AES256-basierend
 - > OpenSSL-Bibliothek

>>> Sicherheit

- * Alle Daten und Metadaten werden signiert - versehentliche und absichtliche Veränderung kann erkannt werden
- * Daten und Metadaten werden verschlüsselt
 - > AES256-basierend
 - > OpenSSL-Bibliothek
 - > Unterstützung für CPU-Beschleunigung (AES-NI, PCLMULQDQ)

>>> Kompression

* Viele Methoden

¹¹aka gzip

>>> Kompression

* Viele Methoden

> zlib¹¹ Gute Kompression bei akzeptabler CPU-Last

¹¹aka gzip

>>> Kompression

* Viele Methoden

- > zlib¹¹ Gute Kompression bei akzeptabler CPU-Last
- > lzma Sehr hohe Kompression bei hoher CPU-Last

¹¹aka gzip

>>> Kompression

* Viele Methoden

- > zlib¹¹ Gute Kompression bei akzeptabler CPU-Last
- > lzma Sehr hohe Kompression bei hoher CPU-Last
- > lz4, lz4hc Speziell für Echtzeitbetrieb, brauchbare Kompression

¹¹aka gzip

>>> Kompression

* Viele Methoden

- > zlib¹¹ Gute Kompression bei akzeptabler CPU-Last
- > lzma Sehr hohe Kompression bei hoher CPU-Last
- > lz4, lz4hc Speziell für Echtzeitbetrieb, brauchbare Kompression

* blosc-Library

- > Hochoptimiert, Hyperthreading, SIMD(AVX2, SSE2)

¹¹aka gzip

>>> Kompression

* Viele Methoden

- > zlib¹¹ Gute Kompression bei akzeptabler CPU-Last
- > lzma Sehr hohe Kompression bei hoher CPU-Last
- > lz4, lz4hc Speziell für Echtzeitbetrieb, brauchbare Kompression

* blosc-Library

- > Hochoptimiert, Hyperthreading, SIMD(AVX2, SSE2)
- > "faster than a memcpy()"

¹¹aka gzip

```
>>> Deduplizierung
```

```
* rolling hash ("buzhash")
```

>>> Deduplizierung

- * rolling hash ("buzhash")
- * Es wird auf Blockebene Dedupliziert -> Auch Dateiteile können erfasst werden

>>> Deduplizierung

- * rolling hash ("buzhash")
- * Es wird auf Blockebene Dedupliziert -> Auch Dateiteile können erfasst werden
 - > Textdateien mit wiederkehrenden Inhalten wie Kopfzeilen oder Textblöcken

>>> Deduplizierung

- * rolling hash ("buzhash")
- * Es wird auf Blockebene Dedupliziert -> Auch Dateiteile können erfasst werden
 - > Textdateien mit wiederkehrenden Inhalten wie Kopfzeilen oder Textblöcken
 - > Betriebssystem in VM-Images

>>> Deduplizierung

- * rolling hash ("buzhash")
- * Es wird auf Blockebene Dedupliziert -> Auch Dateiteile können erfasst werden
 - > Textdateien mit wiederkehrenden Inhalten wie Kopfzeilen oder Textblöcken
 - > Betriebssystem in VM-Images
- * Erfasst auch verschobene/umbenannte Dateien

>>> Deduplizierung

- * rolling hash ("buzhash")
- * Es wird auf Blockebene Dedupliziert -> Auch Dateiteile können erfasst werden
 - > Textdateien mit wiederkehrenden Inhalten wie Kopfzeilen oder Textblöcken
 - > Betriebssystem in VM-Images
- * Erfasst auch verschobene/umbenannte Dateien
- * Auf Repository-Ebene, also auch von mehreren Rechnern o.Ä. nutzbar

>>> Erste Erfahrungen und Ausblick

* 4 VMs, 180GB Daten, 10 Revisionen

>>> Erste Erfahrungen und Ausblick

* 4 VMs, 180GB Daten, 10 Revisionen-> 77GB

>>> Erste Erfahrungen und Ausblick

- * 4 VMs, 180GB Daten, 10 Revisionen-> 77GB
- * ca. 15TB VM-Repo

>>> Erste Erfahrungen und Ausblick

- * 4 VMs, 180GB Daten, 10 Revisionen-> 77GB
- * ca. 15TB VM-Repo
- * Bisher keine Abstürze, bei Abbrüchen keine Fehler

>>> Erste Erfahrungen und Ausblick

- * 4 VMs, 180GB Daten, 10 Revisionen-> 77GB
- * ca. 15TB VM-Repo
- * Bisher keine Abstürze, bei Abbrüchen keine Fehler
- * Webinterface (BorgWeb)