Deduplizierung mit Bacula Base Jobs

Bacula Base Jobs

Who? Philipp Storz

From? http://www.dass-it.de/

When? FrOSCon.de, 26.08.2012

Rev : 17207

dass IT

Philipp Storz

- seit 1998 Beschäftigung mit Linux
- seit 2001 Vollzeit
- seit 2004 - Geschäftsführer dass IT GmbH
- seit 2008 - Jährliche Veranstaltung Bacula Konferenz
- 2012 Bacula Buch - Open Source Press

dass IT GmbH

Eckdaten

- Geschäftsfelder rund um Open Source:
  - Consulting
  - Support
  - Maßgeschneiderte Anpassungen und Entwicklung
- Gegründet: 2004
- Mitarbeiter: 8

dass IT

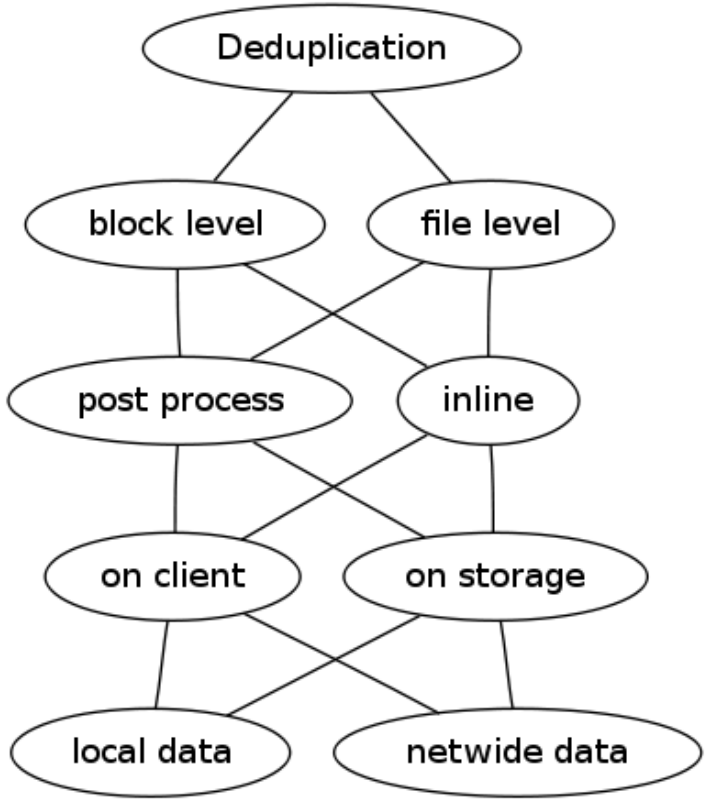
Deduplizierung mit Bacula Base Jobs

Agenda

- Was ist Deduplizierung?
- Was ist Bacula und was ist Accurate Backup?
- Was sind Base Jobs?
- Base Jobs in der Praxis.

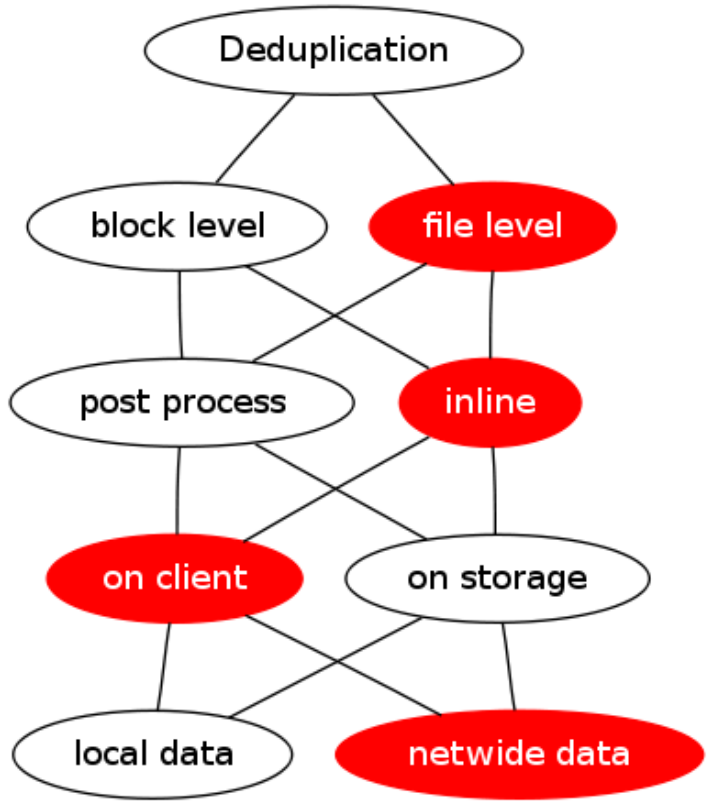
Klassifizierung der Deduplizierungstechniken

- Wie wird dedupliziert?
- Wann wird dedupliziert?
- Wo wird dedupliziert?
- Was wird dedupliziert?



Klassifizierung Bacula Base Jobs

- auf Dateiebene
- während der Sicherung
- auf dem Client
- Daten werden netzweit dedupliziert



Bacula Übersicht

- Open-Source (GPL)
- Projektstart: 2000
- netzwerkbasiert
- skalierbar
- unterstützt alle gängigen Betriebssysteme
- Code Qualität und Dokumentation sind hervorragend

Aufgaben eines Backup Systems

- Lesen und schreiben auf dem Endgerät
- Lesen und schreiben auf Sicherungsmedien
- Sicherungen planen und katalogisieren
- (Rück-)Sicherungen durchzuführen über Anwenderschnittstelle

Daten Lesen und schreiben: File Daemon

- Lesen, Schreiben und Verifizieren von Dateien
- Lesen und Schreiben von ACLs, Attributen
- Ausführen von VSS Snapshots unter Windows
- Berechnung von Prüfsummen
- ...

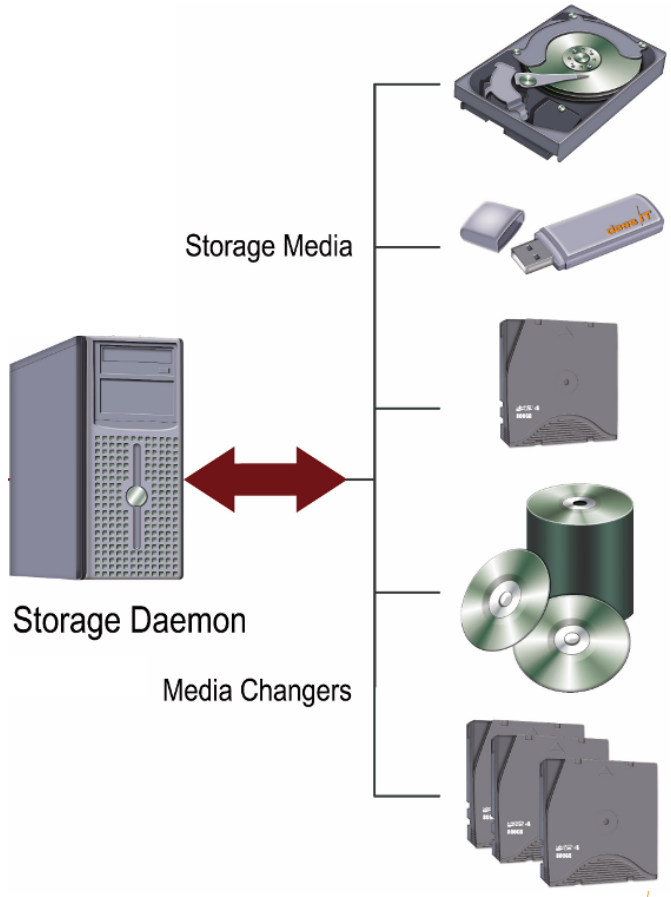
Windows File Daemon

Unix File Daemon

Mac File Daemon

Lesen und schreiben auf Sicherungsmedien: Storage Daemon

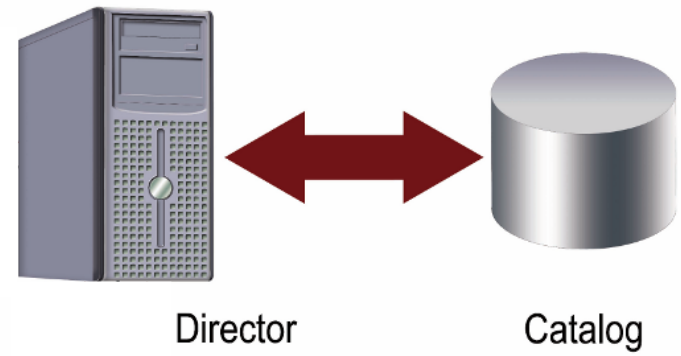
- Ansteuerung von Laufwerken: Disk, Tape, WORM, DVD, etc.
- Ansteuerung von Medienwechslern
- Lesen von physischen Medienlabels (Barcodes)
- Schreiben von logischen Bacula-Labeln auf die Medien
- Durchführung von Kopien, Migrationen und virtuellen Backups
- Medienfehler verarbeiten



dass IT

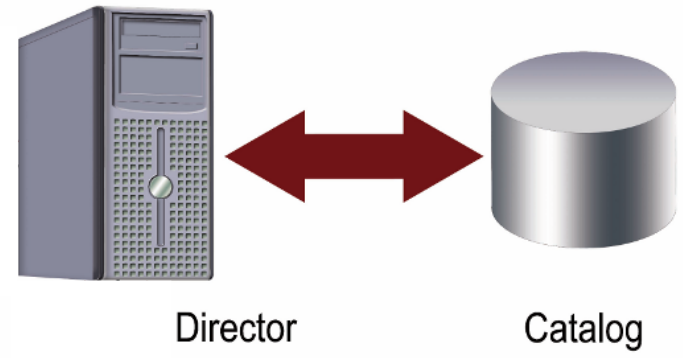
Sicherungen planen und katalogisieren: Bacula Director

- Verwaltung des Bacula Catalog
- Medienverwaltung, Pools
- Zeitplanung
- Konfiguration der zu sichernden Dateien
- Konfiguration der Backupstrategie: voll, inkrementell, differenziell
- Meldungen, Statistiken und Berichte
- Ausführung von Skripten



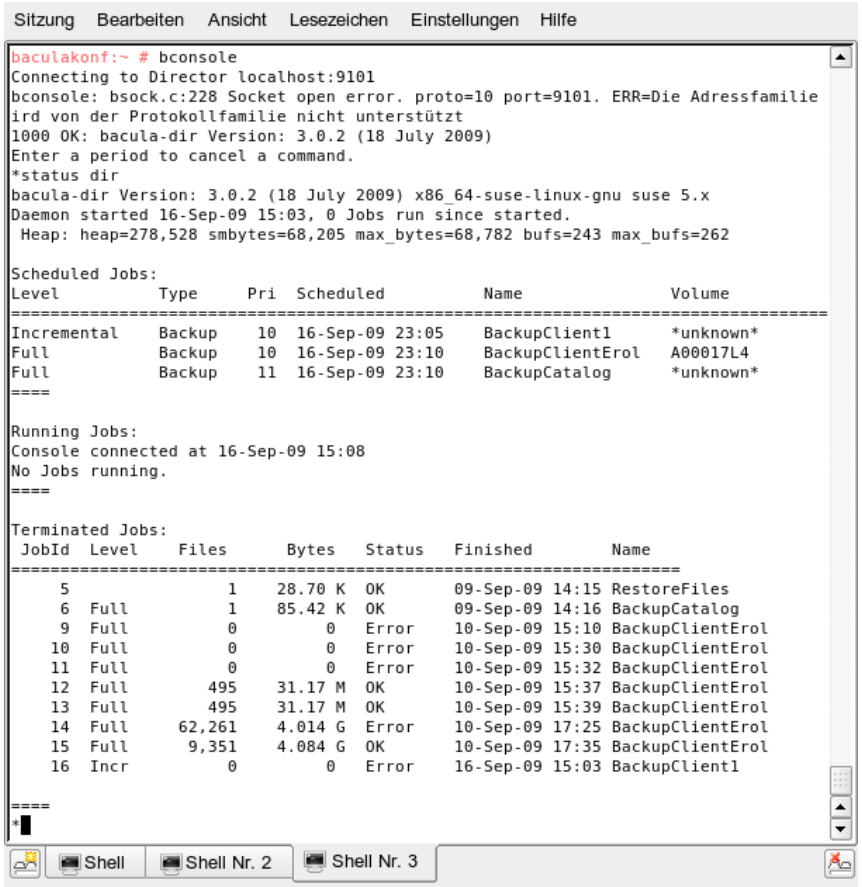
Bacula Catalog

- Speichert Informationen über alle Dateien, Medien und Jobs
- Backend: MySQL,PostgreSQL oder SQLite
- Umfangreiche Suchmöglichkeit vorhanden
- Über sql beliebige Suchen möglich



Bacula Console

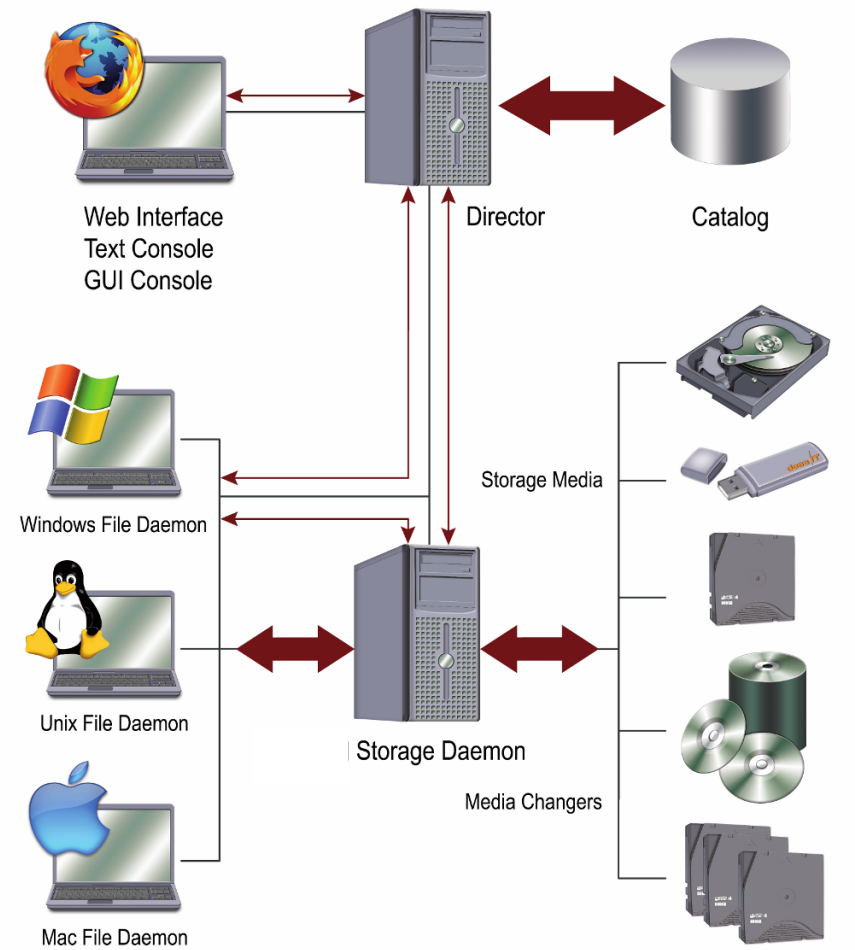
- Anwenderschnittstelle für Rücksicherungen
- Anzeige des Zustands des Systems
- Anzeige des Catalogs
- Starten von Sicherungen



dass IT

Das Gesamtbild

- Kommunikation erfolgt über TCP/IP
- Es werden definierte Ports genutzt
- Kommunikation kann per TLS gesichert werden.



Backuplevel

- Vollbackup
  - großer zeitliche Abstand, z.B. Monat
  - Alles wird gesichert
- Differenzielle Sicherung
  - mittlerer zeitlicher Abstand, z.B. Woche
  - Differenz zur letzten Vollsicherung
- Inkrementelle Sicherung
  - kleiner zeitlicher Abstand, z.B. Tag
  - Differenz zur letzten beliebigen Sicherung

Wie wird die Differenz festgestellt?

- Jede Datei besitzt Zeitstempelinformationen
- Zeitstempel wird bei Veränderung automatisch neu gesetzt
- Referenzsicherung hat einen bestimmten Zeitstempel
- Differenz = allen Dateien, deren Zeitstempel neuer als der Referenzzeitpunkt ist.

Zusammenfassung Zeitstempelbasierte Sicherungstechnik

- Differenziellen / Inkrementelle Sicherung:
  - Es gibt einen Referenzzeitpunkt.
  - Dies ist der Zeitpunkt der vorherige Sicherung.
  - Es werden alle Dateien gesichert, bei denen der Zeitstempel neuer ist als der Referenzzeitpunkt.

Probleme beim Zeitstempelverfahren

Ursache

- Datei wird mit altem Zeitstempel erzeugt
- Dateibaum wird verschoben
- Programm manipuliert Zeitstempel (z.B. Virenscanner)

Wirkung

- Zu sichernde Dateien werden nicht gesichert
- Nicht zu sichernde Dateien werden gesichert
- Gelöschte Dateien werden wiederhergestellt

Lösung: Accurate Backup

- Übertragung der Informationen zu bekannten Dateien und Verzeichnissen an den Client
- Zeitstempelprobleme vermeiden durch Vergleich von:
  - Inodes
  - Zugriffsrechte
  - Anzahl der Links
  - Benutzer und GruppenID
  - Größe
  - Zugriffszeit
  - Modificationszeit Änderungszeit
  - MD5 und SHA1 Signatur.
- Nachteil: Ressourcenintensiv

Verringerung des Backupvolumens

Ink/Diff

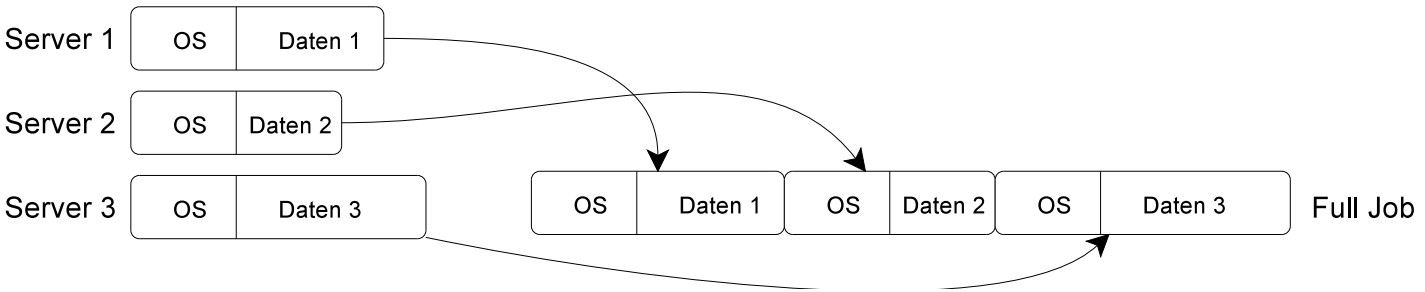
- Alle Dateien werden weggelassen, die bei der Referenzsicherung *dieses Systems* bereits gesichert wurden.

Base Jobs

- Alle Dateien werden weggelassen, die bei der Referenzsicherung *irgendeines Systems* bereits gesichert wurden.

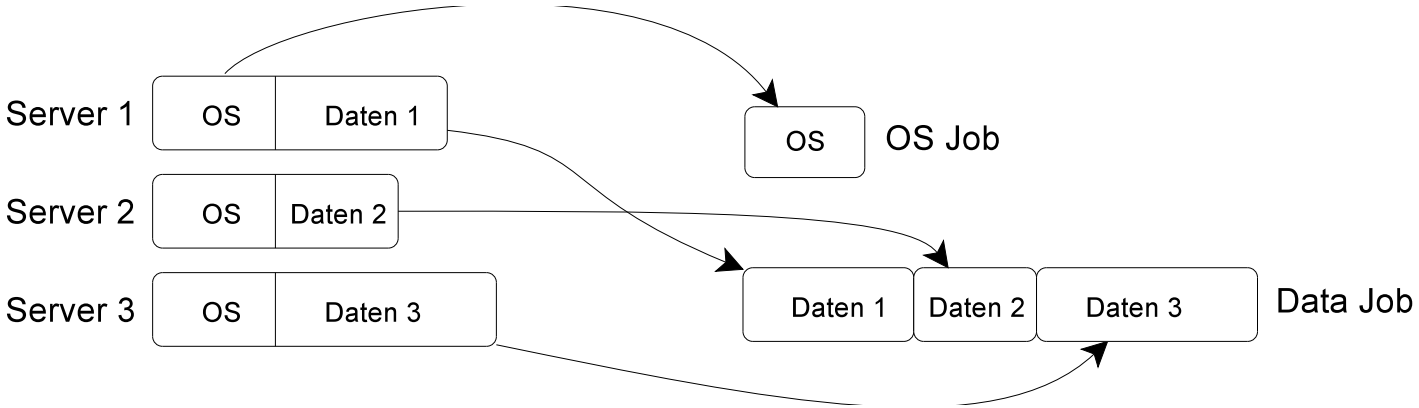
Sicherung identischer Systeme

Bei identischen Systemen werden die Daten und OS mehrfach gesichert.



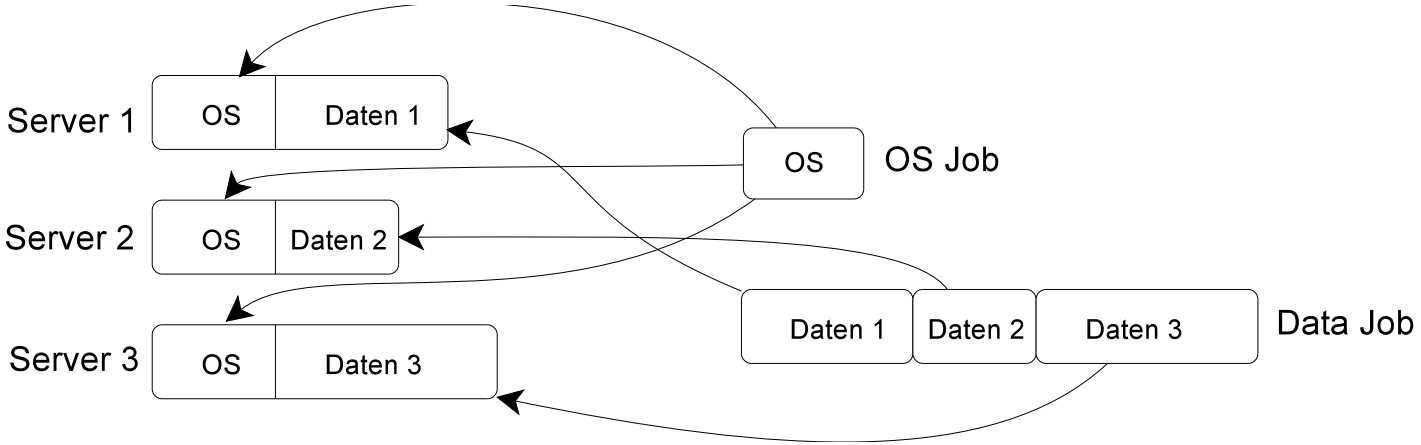
Grundüberlegung bei Basejobs

Bei identischen Systemen wird der gemeinsame Teil nur einmal gesichert.



Rücksicherung bei Basejobs

Bei Rücksicherungen wird der individuelle Teil aus der eignen Sicherung, der gemeinsame Teil aber aus dem Base Jobs zurückgespielt



Ablauf beim Einsatz von Base Jobs

1. Sicherung eines **Base Jobs** in dem möglichst die identischen Daten aller Server gesichert werden.
2. "*Normale*" Sicherungen, bei der zu referenzierende Base Job eingestellt wird. *Accurate* ist notwendig.
3. Wenn die gemeinsame Basis sich erheblich ändert, erneut einen Base Job ausführen.

Praxis: Einsatz von Base Jobs

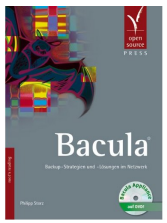
1. VM1: froscon-server
   - director daemon
   - storage daemon
   - file daemon
2. VM2: froscon-client
   - file daemon
3. Jobs
   - froscon-base-job
   - froscon-client-job (basierend auf froscon-base-job)
   - froscon-server-job (basierend auf froscon-base-job)

Zusammenfassung

1. Mit den Einsatz von **Base Jobs** kann das Backupvolumen bei identischen Systemen erheblich verringert werden.
2. Bacula Base Jobs arbeiten auf Dateieebene. Für den Einsatz von Base Jobs ist *Accurate Backup* ist notwendig.
3. Gemeinsame Dateien werden nur einmal gespeichert.
4. Die Deduplizierung erfolgt bereits auf dem Client.

In eigener Sache

1. Diplomanden gesucht im Bereich Open Source / Linux / C / Python
2. Bacula Buch
3. Bacula Konferenz



http://www.dass-it.de
Philipp.Storz@dass-it.de

Fragen?