Das Z-Dateisystem

Speichern - aber sicher!

Jan Schampera <jan.schampera@unix.net>

Städtische Berufschule III
Regensburg

Februar 2008

Gliederung

Massenspeicher
Wie wir Massenspeicher nutzen

„Kleines Nutzungsprofil”

- Massiver Speicherbedarf

Massenspeicher
Wie wir Massenspeicher nutzen

„Kleines Nutzungsprofil”

- Massiver Speicherbedarf
- Täglicher Bedarfszuwachs

Massenspeicher
Wie wir Massenspeicher nutzen

„Kleines Nutzungsprofil”

- Massiver Speicherbedarf
- Täglicher Bedarfszuwachs
- Gefühlte Sicherheit der Hardware

Massenspeicher
Wie wir Massenspeicher nutzen

„Kleines Nutzungsprofil”

- Massiver Speicherbedarf
- Täglicher Bedarfszuwachs
- Gefühlte Sicherheit der Hardware
- Vertrauen der Nutzer in Datensicherheit

Klassisch
Von Volume-Managern und Dateisystemen

Volume-Management

- Erzeugung logischer Datenträger und Anpassung an Gegebenheiten

Klassisch
Von Volume-Managern und Dateisystemen

Volume-Management

- Erzeugung logischer Datenträger und Anpassung an Gegebenheiten
- Auch mehrere physikalische Datenträger

Klassisch
Von Volume-Managern und Dateisystemen

Volume-Management

- Erzeugung logischer Datenträger und Anpassung an Gegebenheiten
- Auch mehrere physikalische Datenträger
- Verschlüsselung und Kompression auf Blockebene

Klassisch
Von Volume-Managern und Dateisystemen

Volume-Management

- Erzeugung logischer Datenträger und Anpassung an Gegebenheiten
- Auch mehrere physikalische Datenträger
- Verschlüsselung und Kompression auf Blockebene
- Software-RAID

Klassisch
Von Volume-Managern und Dateisystemen

Dateisysteme

- Sehr hohe Speicherabstraktionsebene

Klassisch
Von Volume-Managern und Dateisystemen

Dateisysteme

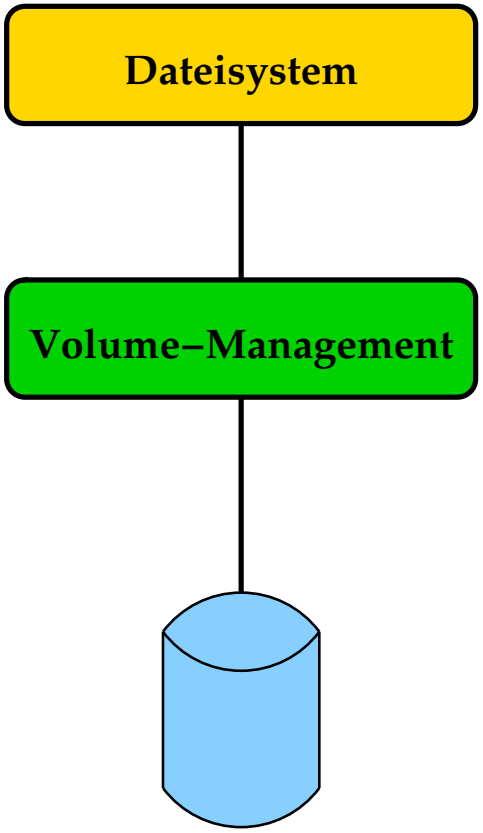
- Sehr hohe Speicherabstraktionsebene
- Zur Organisation von Daten

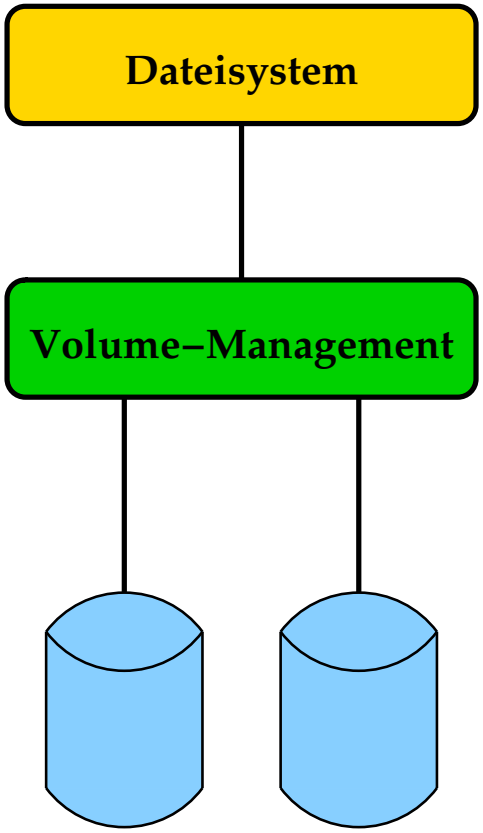
Klassisch
Von Volume-Managern und Dateisystemen

Dateisysteme

- Sehr hohe Speicherabstraktionsebene
- Zur Organisation von Daten
- In verschiedensten Ausprägungen, je nach Anwendungszweck

Klassisch
Von Volume-Managern und Dateisystemen





Klassisch
Von Volume-Managern und Dateisystemen

Veraltete Annahmen

- Datenpfade „sind sicher”

Klassisch
Von Volume-Managern und Dateisystemen

Veraltete Annahmen

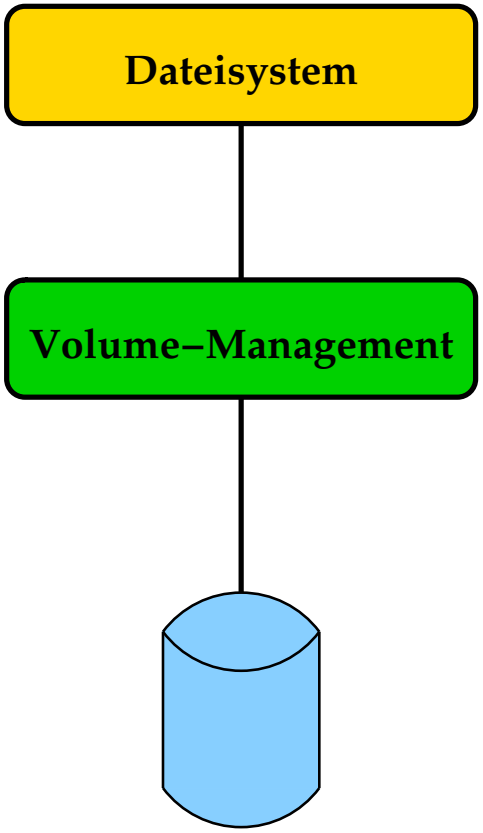
- Datenpfade „sind sicher”
- Manipulation „existiert nicht”

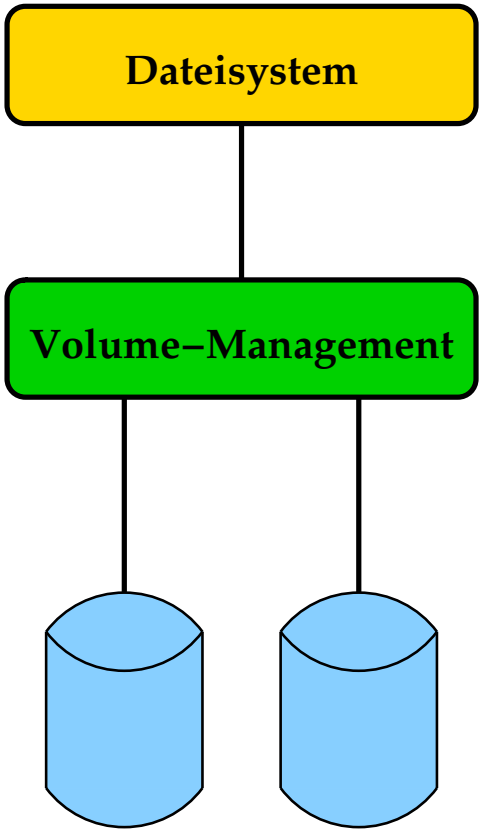
Klassisch
Von Volume-Managern und Dateisystemen

Veraltete Annahmen

- Datenpfade „sind sicher”
- Manipulation „existiert nicht”
- Massenspeicher „halten ewig”

Klassisch
Von Volume-Managern und Dateisystemen





Sicherheitsmaßnahmen
Gängige Vorsicht

Redundanzen bilden

- Erzeugung mehrerer identischer Kopien

Sicherheitsmaßnahmen
Gängige Vorsicht

Redundanzen bilden

- Erzeugung mehrerer identischer Kopien
- Innerhalb der Ausfallzeit trotzdem Daten liefern

Sicherheitsmaßnahmen
Gängige Vorsicht

Redundanzen bilden

- Erzeugung mehrerer identischer Kopien
- Innerhalb der Ausfallzeit trotzdem Daten liefern
- Eventuelle Ersatzmedien sofort wieder beschreiben

Sicherheitsmaßnahmen
Gängige Vorsicht

Blockprüfsummen

- Erzeugung durch Laufwerkselektronik oder Subsystem-Controller

Sicherheitsmaßnahmen
Gängige Vorsicht

Blockprüfsummen

- Erzeugung durch Laufwerkselektronik oder Subsystem-Controller
- Speicherung transparent mit auf dem Datenträger

Sicherheitsmaßnahmen
Gängige Vorsicht

Blockprüfsummen

- Erzeugung durch Laufwerkselektronik oder Subsystem-Controller
- Speicherung transparent mit auf dem Datenträger
- Eindeutige Erkennung von fehlerhaften Blöcken

Sicherheitsmaßnahmen
Gängige Vorsicht

Blockprüfsummen

- Erzeugung durch Laufwerkselektronik oder Subsystem-Controller
- Speicherung transparent mit auf dem Datenträger
- Eindeutige Erkennung von fehlerhaften Blöcken
- Zur Auswahl der „überlebenden Kopie” im Fehlerfall

Sicherheitsmaßnahmen
Gängige Vorsicht

Backups

- Zentralbackups sehr zeitaufwendig, verbrauchen teilweise enorme Ressourcen

Sicherheitsmaßnahmen
Gängige Vorsicht

Backups

- Zentralbackups sehr zeitaufwendig, verbrauchen teilweise enorme Ressourcen
- Volles Recovery oft zu langwierig (veraltete Datenstände!)

Sicherheitsmaßnahmen
Gängige Vorsicht

Backups

- Zentralbackups sehr zeitaufwendig, verbrauchen teilweise enorme Ressourcen
- Volles Recovery oft zu langwierig (veraltete Datenstände!)
- Recovery einzelner Dateien unverhältnismäßig aufwendig

Sicherheitsmaßnahmen
Gängige Vorsicht

Backups

- Zentralbackups sehr zeitaufwendig, verbrauchen teilweise enorme Ressourcen
- Volles Recovery oft zu langwierig (veraltete Datenstände!)
- Recovery einzelner Dateien unverhältnismäßig aufwendig
- Aber natürlich sind Backups durch nichts zu ersetzen!

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Ausfall eines nicht-redundanten Speichers

- Bei Totalausfall: Kompletter Datenverlust

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Ausfall eines nicht-redundanten Speichers

- Bei Totalausfall: Kompletter Datenverlust
- Teilausfälle durch Magnetisierungs- oder Materialfehler

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Ausfall eines nicht-redundanten Speichers

- Bei Totalausfall: Kompletter Datenverlust
- Teilausfälle durch Magnetisierungs- oder Materialfehler
- „Schleichende Zersetzung”

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Ausfall eines nicht-redundanten Speichers

- Bei Totalausfall: Kompletter Datenverlust
- Teilausfälle durch Magnetisierungs- oder Materialfehler
- „Schleichende Zersetzung”
- Stichwort: Lebensdauer

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Bedingte Fehlererkennung

- Bei Datenspiegelung unmöglich die „richtige” Spiegelhälfte zu erkennen

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Bedingte Fehlererkennung

- Bei Datenspiegelung unmöglich die „richtige” Spiegelhälfte zu erkennen
- Eine Lösung: Transparente Prüfsummen durch Subsystem-Controller abgespeichert

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Bedingte Fehlererkennung

- Bei Datenspiegelung unmöglich die „richtige” Spiegelhälfte zu erkennen
- Eine Lösung: Transparente Prüfsummen durch Subsystem-Controller abgespeichert
- In jedem Fall ist das Betriebssystem bei der Prüfung „außen vor”

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Unsichere Datenpfade

- Falsches Denken: Controller „funktionieren zu 100% oder garnicht”

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Unsichere Datenpfade

- Falsches Denken: Controller „funktionieren zu 100% oder garnicht”
- Fremdlicht (Fibre Channel) oder Fremdspannungen

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Unsichere Datenpfade

- Falsches Denken: Controller „funktionieren zu 100% oder garnicht”
- Fremdlicht (Fibre Channel) oder Fremdspannungen
- Firmwarefehler, Hardwarefehler

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Unsichere Datenpfade

- Falsches Denken: Controller „funktionieren zu 100% oder garnicht”
- Fremdlicht (Fibre Channel) oder Fremdspannungen
- Firmwarefehler, Hardwarefehler
- Magnetfelder, Streustrahlung

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Unsichere Datenpfade

- Falsches Denken: Controller „funktionieren zu 100% oder garnicht”
- Fremdlicht (Fibre Channel) oder Fremdspannungen
- Firmwarefehler, Hardwarefehler
- Magnetfelder, Streustrahlung
- Gewollte Manipulation

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Schadcode

- Computerviren „verstecken sich” oft in ungenutzten Bereichen von Dateien oder des Dateisystems

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Schadcode

- Computerviren „verstecken sich” oft in ungenutzten Bereichen von Dateien oder des Dateisystems
- Dateisysteme erkennen „Eindringlinge” nicht mittels ihrer Organisationsstruktur

Gefährdung der Daten

Alltägliche und nicht ganz alltägliche Probleme

Schadcode

- Computerviren „verstecken sich” oft in ungenutzten Bereichen von Dateien oder des Dateisystems
- Dateisysteme erkennen „Eindringlinge” nicht mittels ihrer Organisationsstruktur
- Neben dem eigentlichen Viruseffekt auch Gefährdung der Nutzdaten durch die Infektion

Das Z-Dateisystem
Eine Lösung für viele Probleme

Kleiner Steckbrief

- Originalname: *„Zettabyte Filesystem”* („ZFS”)

Das Z-Dateisystem
Eine Lösung für viele Probleme

Kleiner Steckbrief

- Originalname: *„Zettabyte Filesystem”* („ZFS”)
- Entwickelt von Sun Microsystems Inc. für Solaris 10

Das Z-Dateisystem
Eine Lösung für viele Probleme

Kleiner Steckbrief

- Originalname: *„Zettabyte Filesystem”* („ZFS”)
- Entwickelt von Sun Microsystems Inc. für Solaris 10
- Kapazität reicht „für immer”:
  $256643977944357425643520 \cdot 10^{21}$ TB pro System

Das Z-Dateisystem
Eine Lösung für viele Probleme

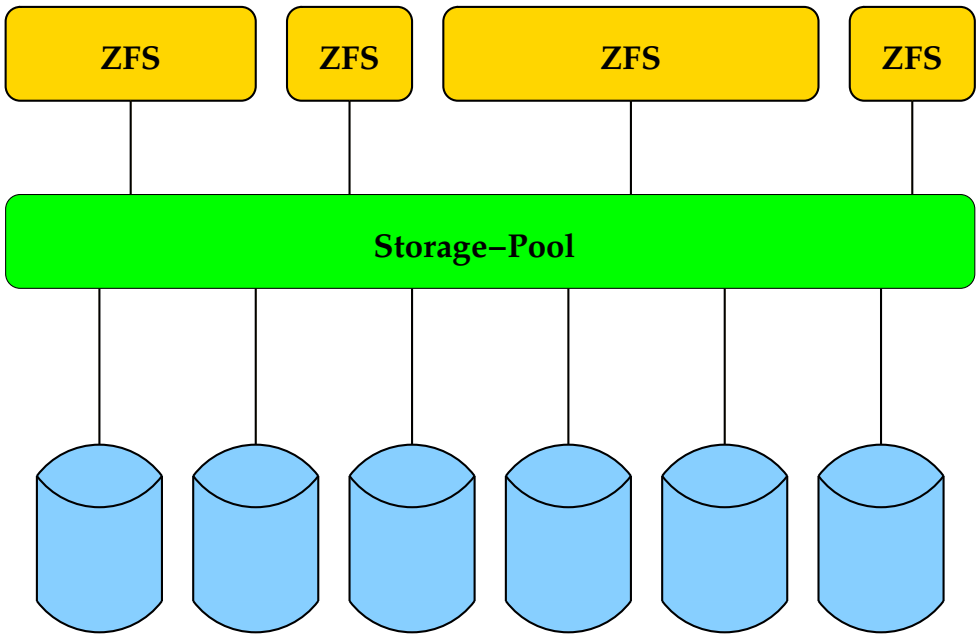
Kleiner Steckbrief

- Originalname: *„Zettabyte Filesystem”* („ZFS”)
- Entwickelt von Sun Microsystems Inc. für Solaris 10
- Kapazität reicht „für immer”:
  $56668397794435742564352 00 \cdot 10^{21}$ TB pro System
- Sehr gute Eigenschaften bzgl. Datenabsicherung

Das Z-Dateisystem
Grundaufbau: Schema

Aufbauschema

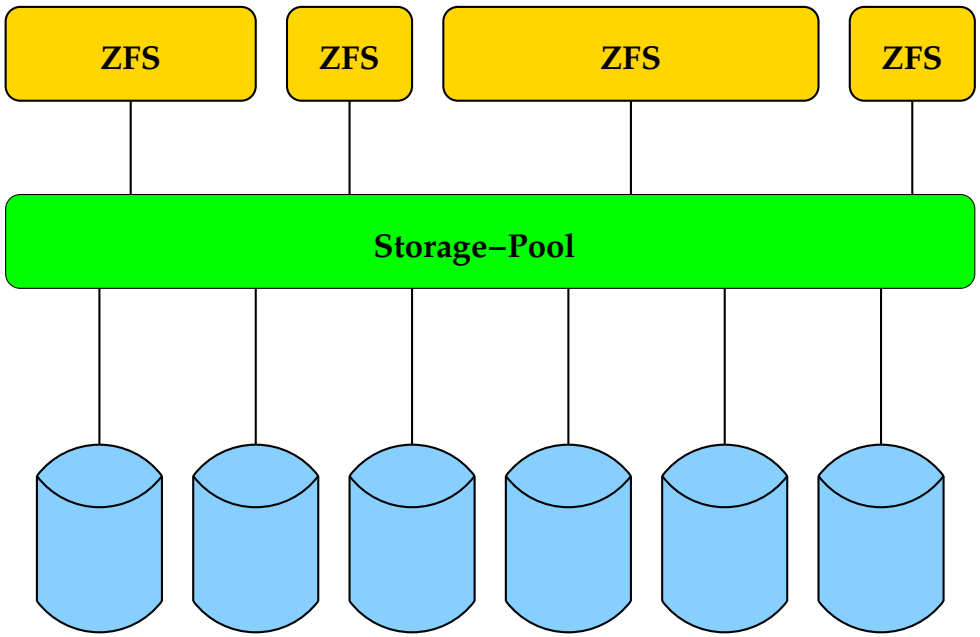
- Datenträger Teil eines Pools (ZPool)



Das Z-Dateisystem
Grundaufbau: Schema

Aufbauschema

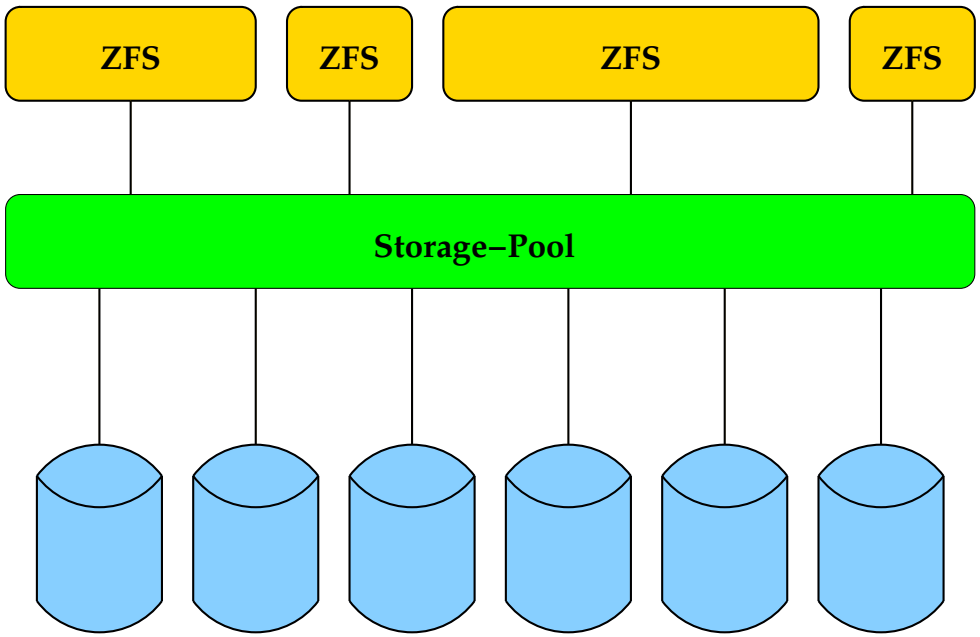
- Datenträger Teil eines Pools (ZPool)
- Blöcke im ZPool redundant und abgesichert



Das Z-Dateisystem
Grundaufbau: Schema

Aufbauschema

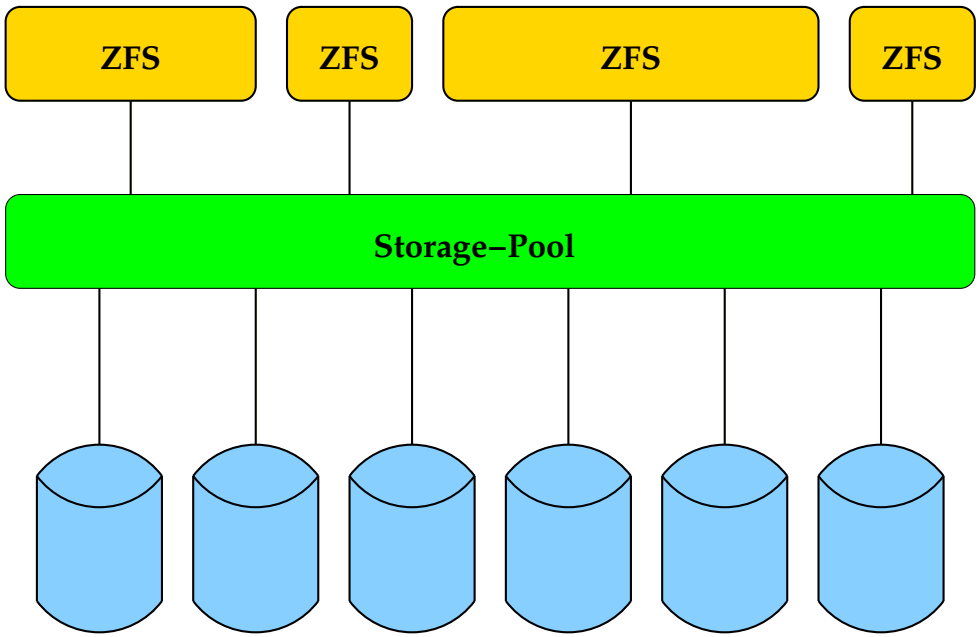
- Datenträger Teil eines Pools (ZPool)
- Blöcke im ZPool redundant und abgesichert
- Alle ZFS „bedienen sich” aus ZPool



Das Z-Dateisystem
Grundaufbau: Schema

Aufbauschema

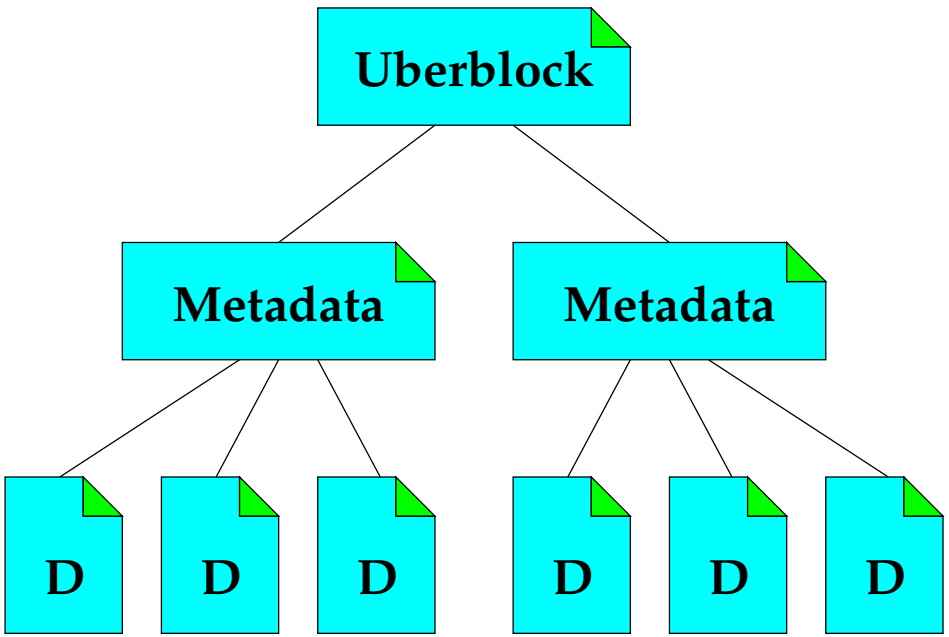
- Datenträger Teil eines Pools (ZPool)
- Blöcke im ZPool redundant und abgesichert
- Alle ZFS „bedienen sich” aus ZPool
- vgl.: Applikationen und RAM



Das Z-Dateisystem
Grundaufbau: Dateisystemorganisation

Dateisystemorganisation

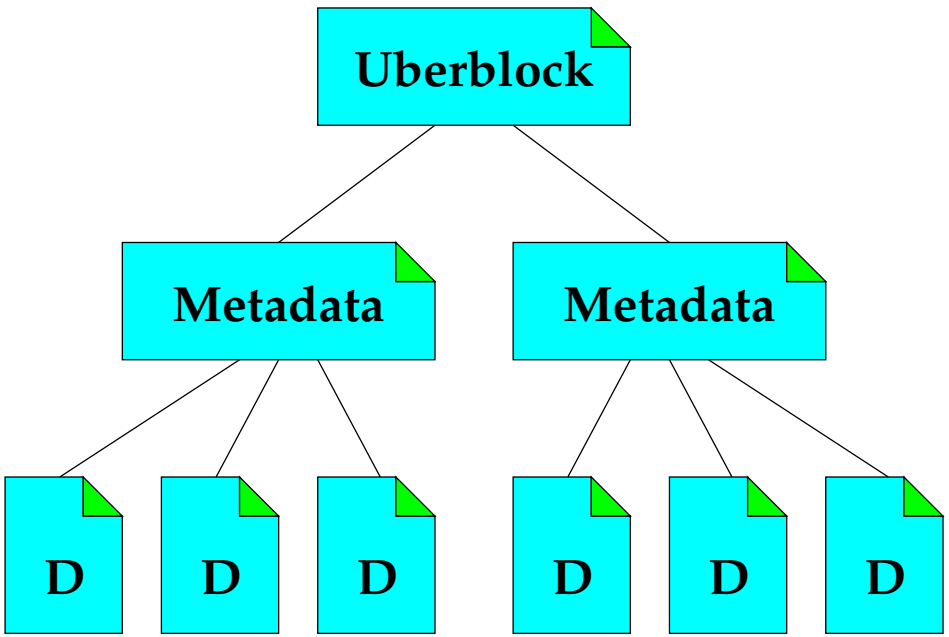
- Baumartige Struktur der Meta- und Nutzdatenblöcke



Das Z-Dateisystem
Grundaufbau: Dateisystemorganisation

Dateisystemorganisation

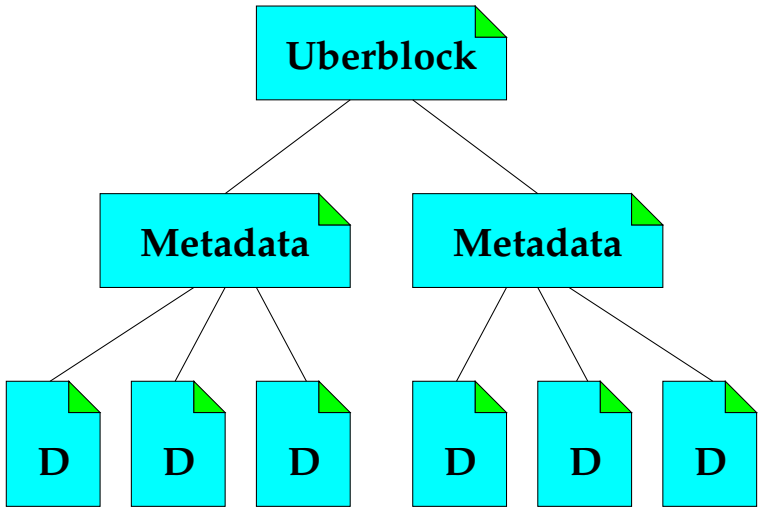
- Baumartige Struktur der Meta- und Nutzdatenblöcke
- Alle Blöcke allokiert im ZPool und somit abgesichert



Das Z-Dateisystem
Absicherung: Datenbankdenken

Transaktionsorientierung

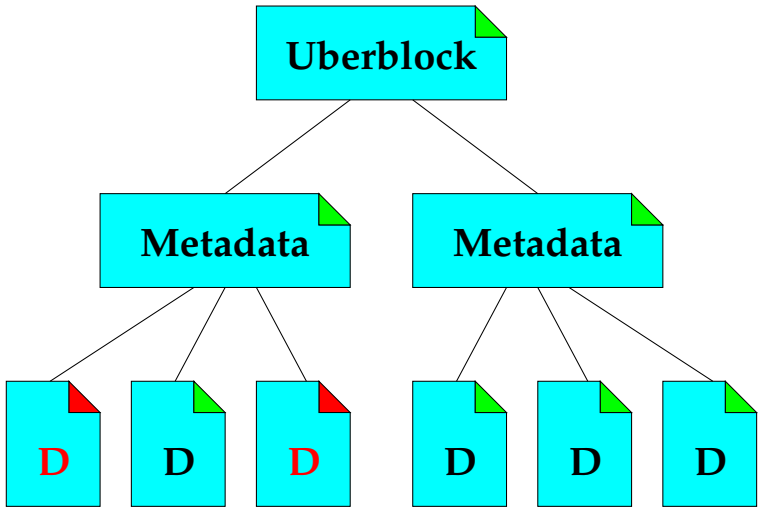
- Schreiben der Blöcke jeder Ebene



Das Z-Dateisystem
Absicherung: Datenbankdenken

Transaktionsorientierung

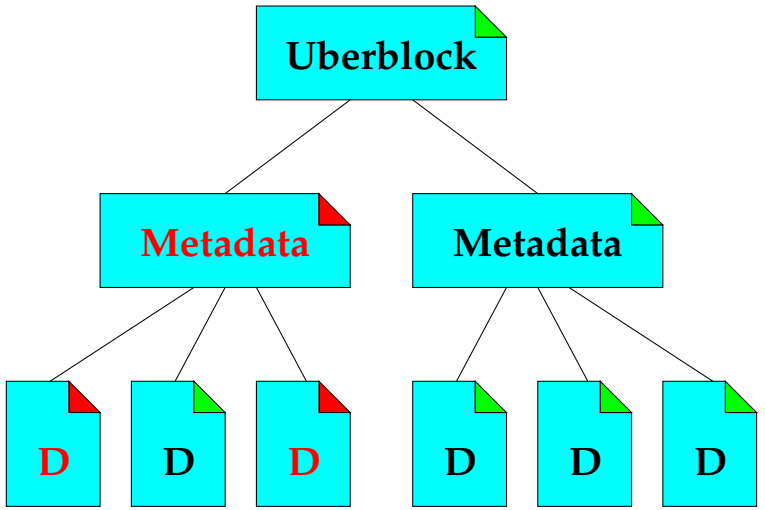
- Schreiben der Blöcke jeder Ebene



Das Z-Dateisystem
Absicherung: Datenbankdenken

Transaktionsorientierung

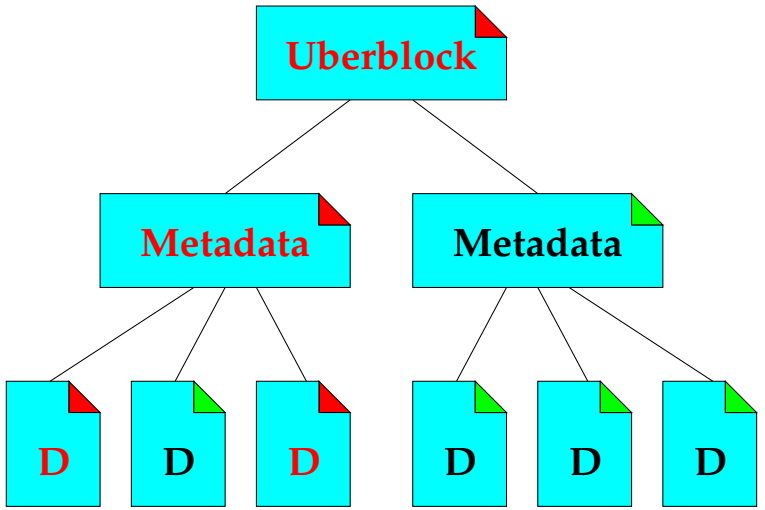
- Schreiben der Blöcke jeder Ebene



Das Z-Dateisystem
Absicherung: Datenbankdenken

Transaktionsorientierung

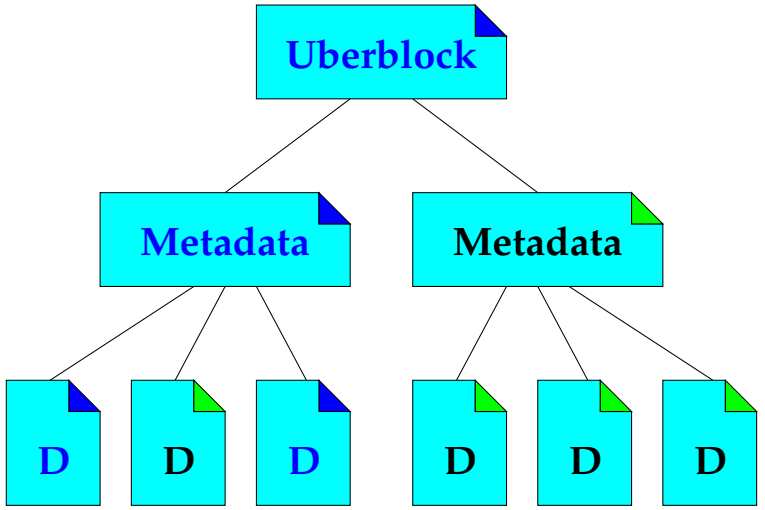
- Schreiben der Blöcke jeder Ebene
- Aktualisierung des „Uberblocks”



Das Z-Dateisystem
Absicherung: Datenbankdenken

Transaktionsorientierung

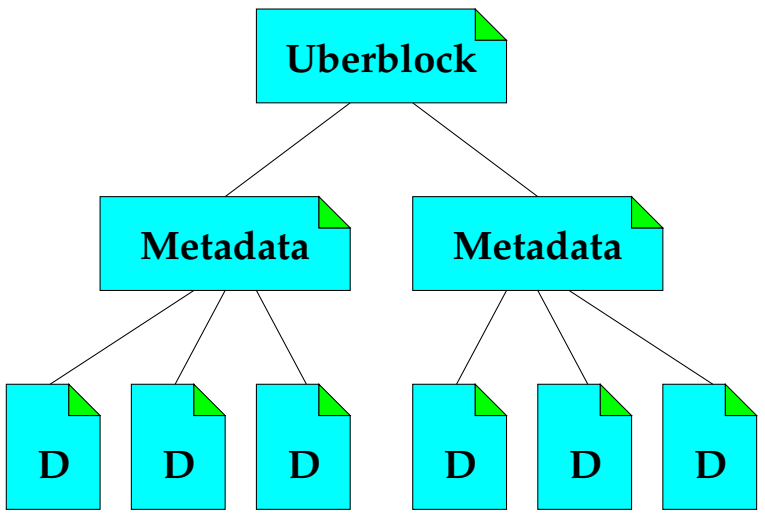
- Schreiben der Blöcke jeder Ebene
- Aktualisierung des „Uberblocks”
- Abschluss der Transaktion



Das Z-Dateisystem
Absicherung: Datenbankdenken

Transaktionsorientierung

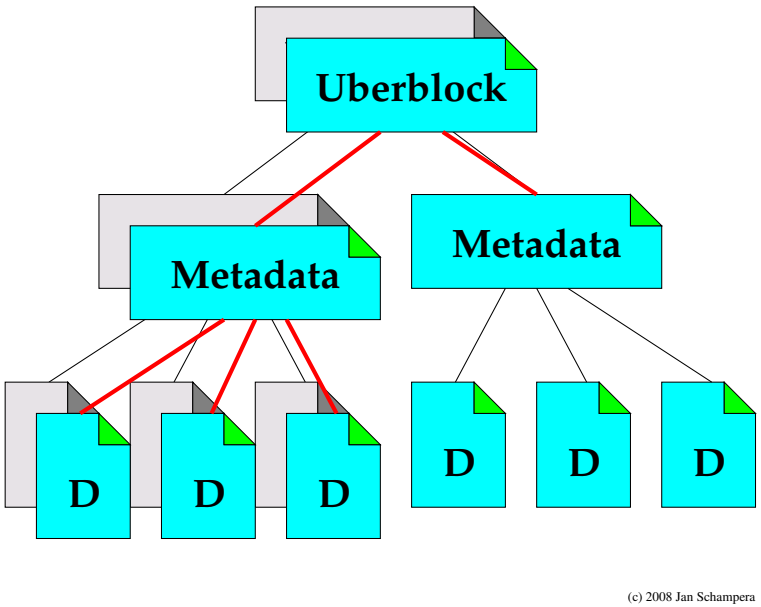
- Schreiben der Blöcke jeder Ebene
- Aktualisierung des „Uberblocks”
- Abschluss der Transaktion
- „Neuer Stand” erreicht



Das Z-Dateisystem
Copy-on-Write

CoW-Semantik

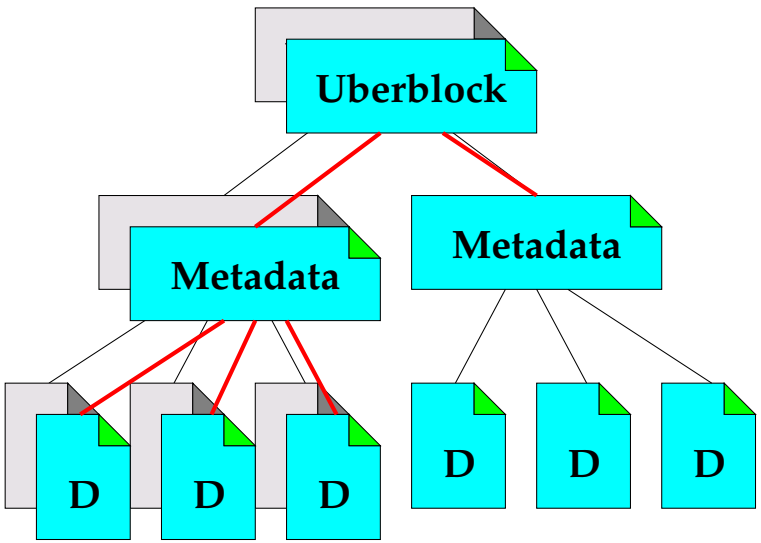
- Schreiboperationen allokieren „frische” Blöcke



Das Z-Dateisystem
Copy-on-Write

CoW-Semantik

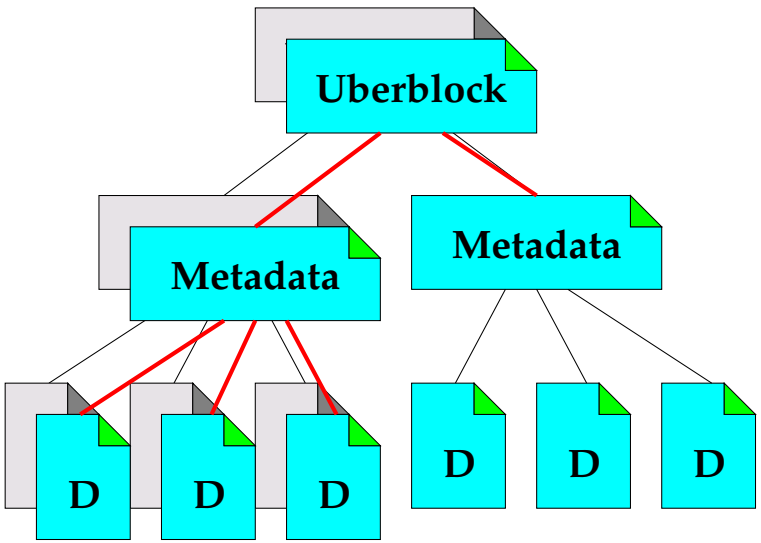
- Schreiboperationen allokieren „frische” Blöcke
- „Alte” Blöcke aufheben



Das Z-Dateisystem
Copy-on-Write

CoW-Semantik

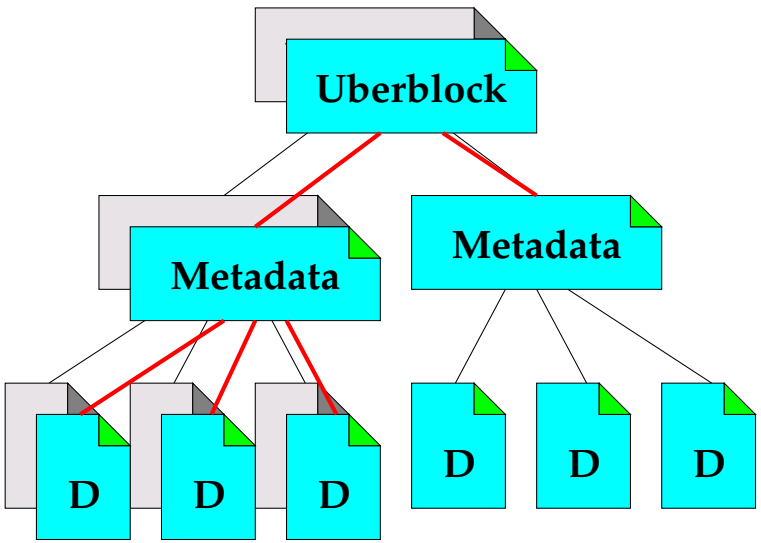
- Schreiboperationen allokieren „frische” Blöcke
- „Alte” Blöcke aufheben
- Anullierung von Transaktionen



Das Z-Dateisystem
Copy-on-Write

CoW-Semantik

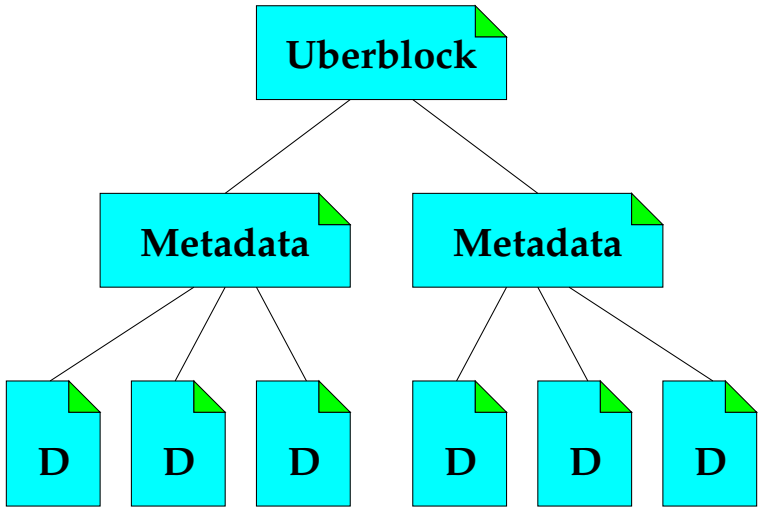
- Schreiboperationen allokieren „frische” Blöcke
- „Alte” Blöcke aufheben
- Anullierung von Transaktionen
- Snapshots



Das Z-Dateisystem
Prüfsummen (Checksums)

Prüfsummen (Checksums)

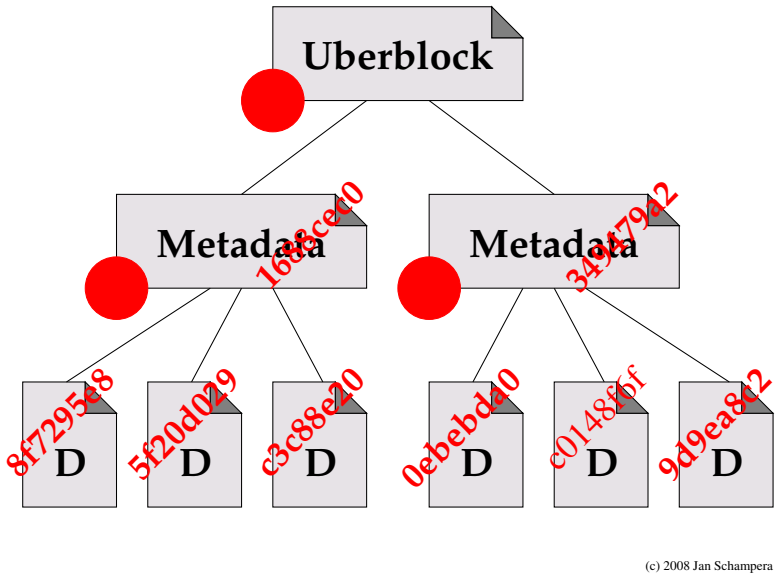
- Prüfsumme über jeden Block



Das Z-Dateisystem
Prüfsummen (Checksums)

Prüfsummen (Checksums)

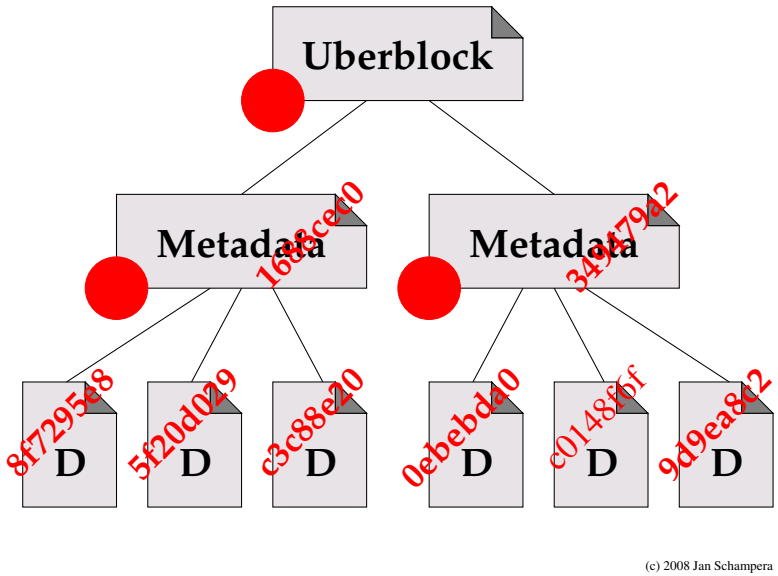
- Prüfsumme über jeden Block
- Speicherung der Prüfsummen jeweils eine Ebene höher



Das Z-Dateisystem
Prüfsummen (Checksums)

Prüfsummen (Checksums)

- Prüfsumme über jeden Block
- Speicherung der Prüfsummen jeweils eine Ebene höher
- Somit komplette Baumstruktur abgesichert



Das Z-Dateisystem

Weitere Features, nicht oder nicht direkt zur Datensicherheit

Einige weitere Features

- Extrem skalierbar

Das Z-Dateisystem

Weitere Features, nicht oder nicht direkt zur Datensicherheit

Einige weitere Features

- Extrem skalierbar
- Bereitstellung von virtuellen Blockgeräten aus ZPool

Das Z-Dateisystem

Weitere Features, nicht oder nicht direkt zur Datensicherheit

Einige weitere Features

- Extrem skalierbar
- Bereitstellung von virtuellen Blockgeräten aus ZPool
- Serialisierte Backupmöglichkeit

Das Z-Dateisystem

Weitere Features, nicht oder nicht direkt zur Datensicherheit

Einige weitere Features

- Extrem skalierbar
- Bereitstellung von virtuellen Blockgeräten aus ZPool
- Serialisierte Backupmöglichkeit
- Erzeugung von Clones aus Snapshotständen

Das Z-Dateisystem

Weitere Features, nicht oder nicht direkt zur Datensicherheit

Einige weitere Features

- Extrem skalierbar
- Bereitstellung von virtuellen Blockgeräten aus ZPool
- Serialisierte Backupmöglichkeit
- Erzeugung von Clones aus Snapshotständen
- Kompression (dateiweise oder kpl. Dateisystem)

ENDE
EOF

Fragen?