

Btrfs



Linux-Dateisystem der Zukunft ?

Inhaltsverzeichnis

- Inhaltsverzeichnis
- Vorgeschichte Btrfs
 - Ziele
- Funktionsumfang
 - Erweiterter Speicherbereich
 - Snapshots und Subvolumes
 - Vergleich
 - Erste Betriebssysteme / Kernel implementierung
- Technischer Aufbau
 - B⁺-Bäume
 - Aufbau: Normales Dateisystem vs. Btrfs Dateisystem
 - Benchmark
- Zukunftsmusik
- Fazit
- Quellen

Vorgeschichte Btrfs

- Btrfs = B-Tree File System *oder* Butter FS
- Analogon zu ZFS
- Idee stammt vom IBM-Forscher Ohad Rodeh
- Copy-on-Write-B-Baum-Struktur
- Wird seit 2007 von Chris Mason (Oracle) entwickelt
- bis heute noch offiziell in der Entwicklungsphase
- Freigabe daher nur für Benchmarks und Reviews
- Btrfs soll sich vom Linux-Standard Dateisystemen mit besonderen Funktionen abheben



Vorgeschichte Btrfs - Ziele

- Allzweck-Dateisystem
- Umgang mit sehr großen Speicherbereichen
- Funktionen, die andere Linux Dateisysteme nicht bieten
- Administrator freundlich, einfach zu handhaben und wenig Fehleranfällig
- Hohe Performance bei variierender Auslastung

Funktionsumfang

- **Erweiterter Speicherbereich von 2^{64} Byte**
- Effizientes Speicher von kleinen Dateien / Ordnern
- Dynamische Inodes
- **Snapshots und mehrere Subvolumes**
- Integriertes Raid
- Checksummen für Daten und Metadaten
- Offline / Online Dateisystemüberprüfung
- Inkrementelles Backup

Funktionsumfang

- **Erweiterter Speicherbereich von 2^{64} Byte**

→ 16 Exabyte oder 16.384 Petabyte oder 16.777.216 Terabyte

Beispiel: ca. 5,59 Mio Festplatten à 3TB Größe

Gestapelt: Turmhöhe 167,7km

Ext4 maximal 16 Terabyte

LHC/CERN 20 Petabyte in einem Peta-Grid

Funktionsumfang

- **Erweiterter Speicherbereich von 2^{64} Byte**
- Effizientes Speicher von kleinen Dateien / Ordnern
- Dynamische Inodes
- **Snapshots und mehrere Subvolumes**
- Integriertes Raid
- Checksummen für Daten und Metadaten
- Offline / Online Dateisystemüberprüfung
- Inkrementelles Backup

Funktionsumfang

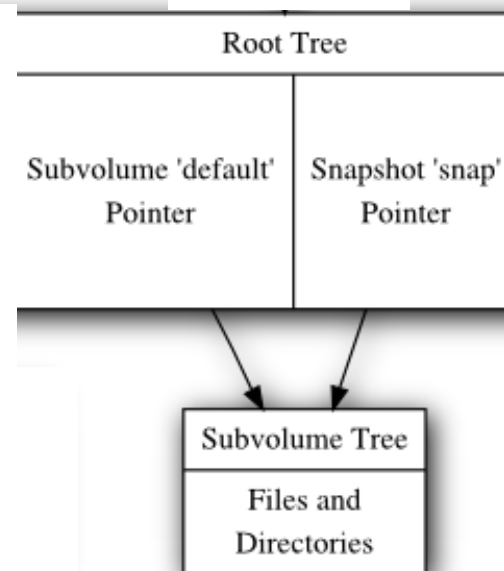
- **Snapshots und mehrere Subvolumes**

- Erstellt „eingefrorene“ Abbilder von einzelnen Dateien, Verzeichnissen oder ganzen Subvolumes
- Erstellung in Bruchteilen von Sekunden
- Snapshots belegen kaum Speicherplatz
- Snapshot verweist auf dieselben Datenblöcke
- Registriert die Veränderungen im Original
- Speichert nur Veränderungen die im Original oder Snapshot geändert wurden

Funktionsumfang

- **Snapshots und mehrere Subvolumes**

- Subvolumes sind eigene Dateisysteme im „Mutter“-Dateisystem



Funktionsumfang

- **Erweiterter Speicherbereich von 2^{64} Byte**
- Effizientes Speicher von kleinen Dateien / Ordnern
- Dynamische Inodes
- **Snapshots und mehrere Subvolumes**
- Integriertes Raid
- Checksummen für Daten und Metadaten
- Offline / Online Dateisystemüberprüfung
- Inkrementelles Backup

Funktionsumfang - Vergleich

	max. Volume	Snapshots	Encryption	Subvolumes	Check
Btrfs	16 EiB	Ja	Geplant	Ja	Ja
Ext3	8 TiB	Nein	Nein	Nein	Nein
Ext4	16 TiB *	Nein	Nein	Nein	Nein
XFS	8 EiB	Ja	Nein	Nein	Nein
ZFS	16 EiB	Ja	Ja	Ja	Ja

* Theoretisch bis 1 EiB, jedoch Beschränkt durch e2fsprogs auf 16 TiB.

Erste Betriebssysteme / Kernel

- Seit 2010 als Standard-Dateisystem bei MeeGo bereits im Einsatz
- Fedora 11 bietet Btrfs als Zusatz-Dateisystem mit an
- Februar 2011: Btrfs geplant als Standard-Dateisystem in Fedora 16
- Ubuntu plant seit Mitte 2010 Btrfs als Standard-Dateisystem zu nutzen; wartet jedoch auf „stable“-Status.
- Im Linux-Kernel seit 2.6.29-rc1 Verfügbar

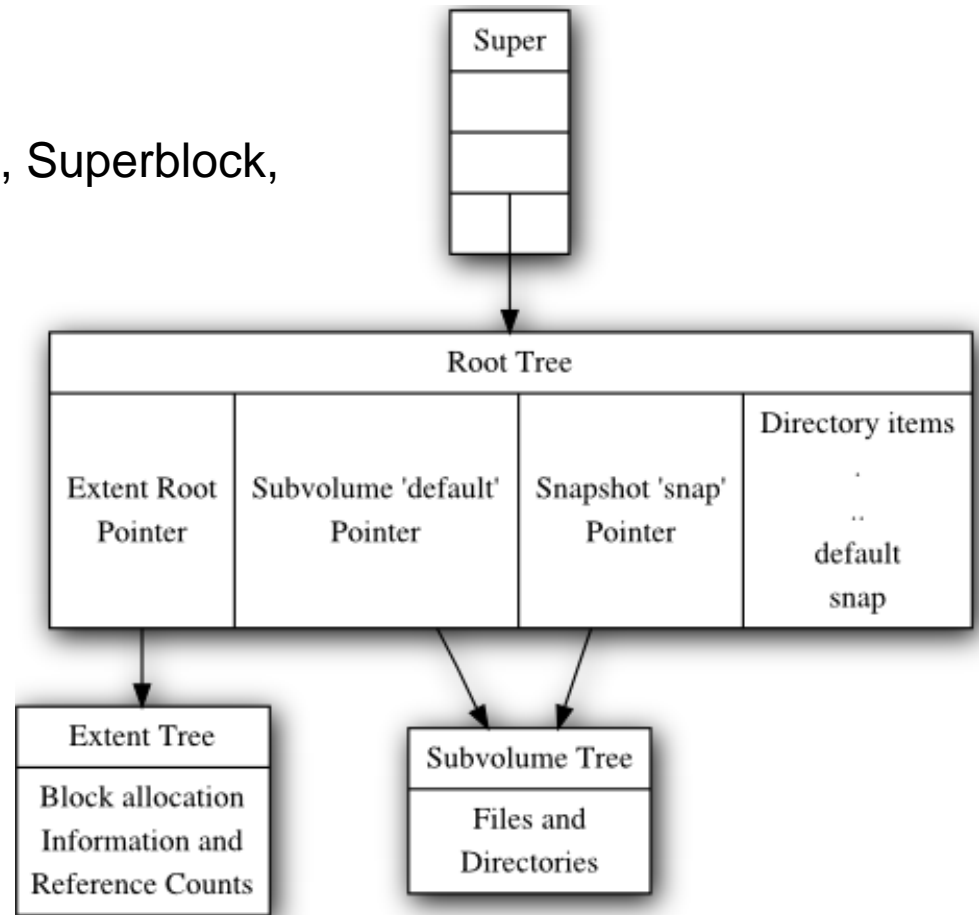
MeeGo™

fedora^f

 ubuntu

Technischer Aufbau

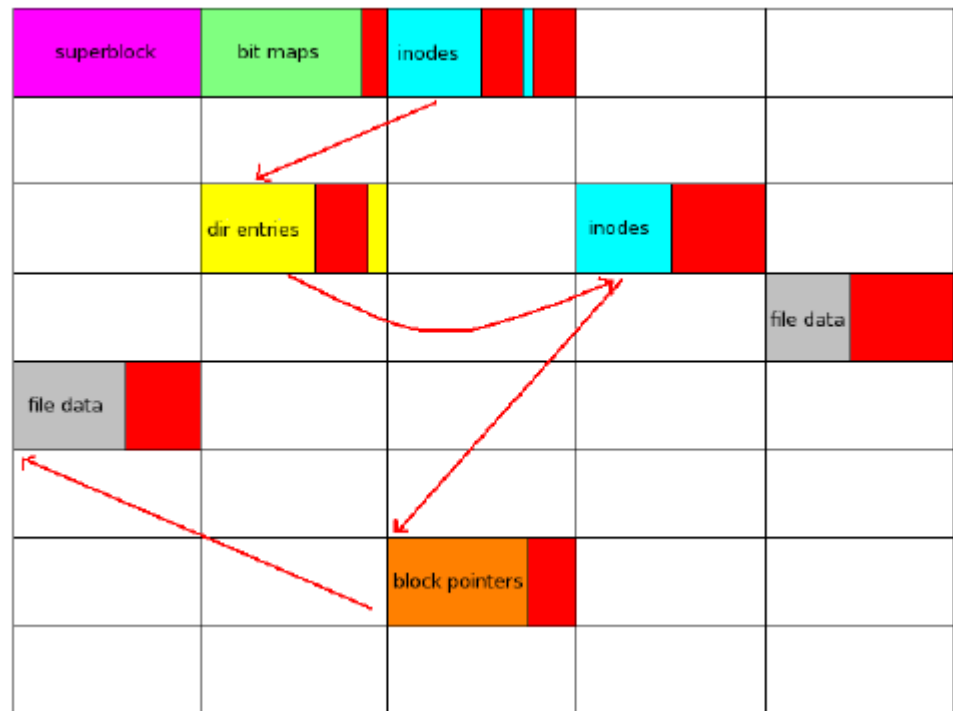
- Btrfs nutzt viele B⁺-Bäume
- Eigene Bäume für Subvolumes, Superblock, Root, Metadaten, etc.
- Bäume verweisen auf andere Bäume



Technischer Aufbau

Normale Dateisysteme:

- Speichern nacheinander in freie Systemblöcke
- Ungenutzter Speicher im Systemblock kann nicht genutzt werden (ROT)
- Lange Suchfolge, um von Metadaten auf die richtigen Daten zu stoßen (ROT Pfeile)



Technischer Aufbau

Btrfs Dateisysteme:

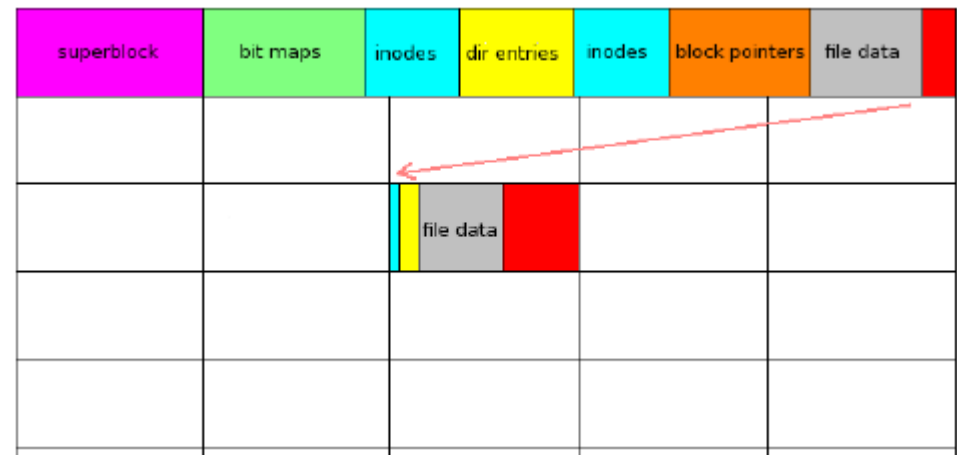
- Alles wird zusammen gepackt
- Sortierung nach Object-ID und Type

<1, Order>

- Kurze Suchfolge, da passende Objekte nebeneinander liegen

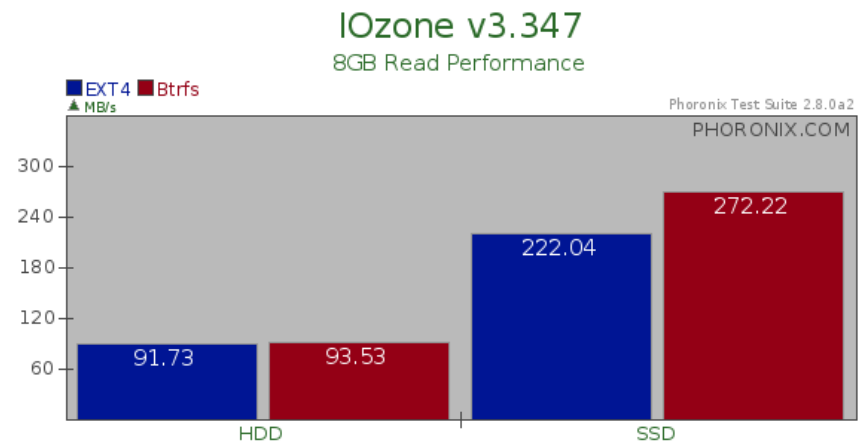
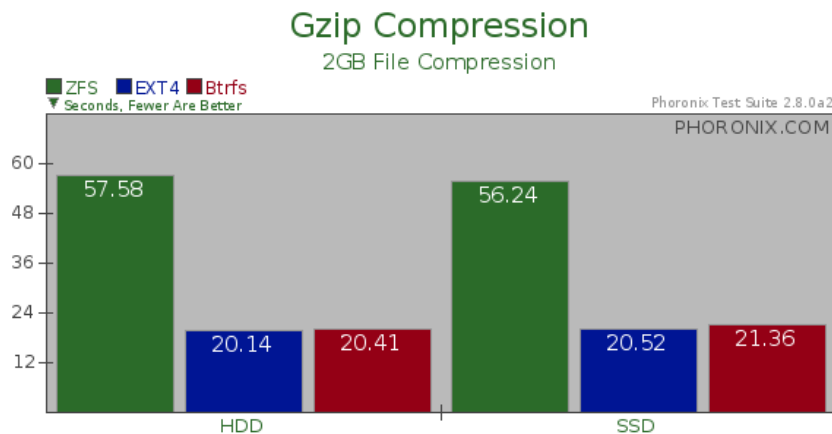
Order „Test“

Datei „A“ aus Order „Test“



Benchmark

- Variierende Ergebnisse zwischen Btrfs und EXT4 im Test
- Je nach Kernel-Version und Btrfs Version
- Ergebnisse von HDD vs. SSD liegen weit auseinander



- Btrfs ist noch nicht das Performance-Wunder wie versprochen
- Jedoch mit sehr guter Tendenz

Zukunftsmusik

- Januar 2011 war Btrfs 1.0 „Stable“-Release geplant
- Keine konkrete Angabe, wann 1.0 Release veröffentlicht wird
 - Vermutungen: März – Mai 2011
- Ubuntu und Fedora planen den Einsatz Btrfs als Standard-Dateisystem
- Btrfs soll irgendwann Ext4 Ablösen
- Netzwerkprotokoll CRFS auf Btrfs aufbauend

Fazit

Linux-Dateisystem der Zukunft?

Ja!

- Besondere Funktionen die bisher gefehlt haben
- Könnte mittelfristig die Standard-Dateisysteme ablösen
- Warten auf 1.0 Release + konkrete Benchmarks
 - Sofern gute Ergebnisse messbar sind, lohnt der zukünftige Einsatz

Quellen

- <http://dclug.tux.org/200908/BTRFS-DCLUG.pdf>
- <http://lwn.net/Articles/342892/>
- <http://www.heise.de/open/meldung/MeeGo-Projekt-waehlt-Btrfs-zum-Standard-Dateisystem-998662.html>
- <http://de.wikipedia.org/wiki/Btrfs>
- <https://btrfs.wiki.kernel.org/index.php/>
- <https://help.ubuntu.com/community/btrfs>
- <http://event.on24.com/event/23/31/61/rt/1/documents/slidepdf/btrfsweb.pdf>
- <http://www.pro-linux.de/artikel/2/1456/btrfs-linux-dateisystem-der-zukunft.html>
- Div. Benchmarks von <http://www.phoronix.com>
- <http://www.heise.de/open/artikel/Snapshots-Subvolumes-Performance-224650.html>
- http://en.wikipedia.org/wiki/Comparison_of_file_systems
- <http://www.thevarguy.com/2010/08/02/ubuntu-10-10s-new-file-system-btrfs/>
- <http://oss.oracle.com/projects/crfs/>

Vielen Dank für eure Aufmerksamkeit



Fragen?