

HAMMER Filesystem

Ein Vortrag von

Timm Müller

Am

13.07.2012

Inhaltsverzeichnis

- Allgemein
 - Was ist HAMMER überhaupt?
 - DragonFlyBSD
 - Was kann HAMMER denn so?
- Features genauer betrachtet
 - Spezifikationen
 - B+ Tree
 - Instant crash recovery
 - Historie
 - Data Deduplication
 - PFSs
 - Mirroring
 - Multi-Volume
- Vergleich
- Fazit

Was ist HAMMER überhaupt?

- 64-bit Cluster-Filesystem
- Standard unter DragonFlyBSD
 - Seit Version 2.0 (2008)
 - Ersetzt UFS
- Ausgelegt für heutige Festplattengrößen

DragonFlyBSD

- Fork von FreeBSD (2003/2004)
- Besonderer Fokus auf Multiprozessorsysteme und Cluster
- Geschaffen von Matthew Dillon
 - Unzufrieden mit Umsetzung von Threading und Multiprocessing
 - FreeBSD nahm seine Ideen nicht an → Fork

Was kann HAMMER denn so?

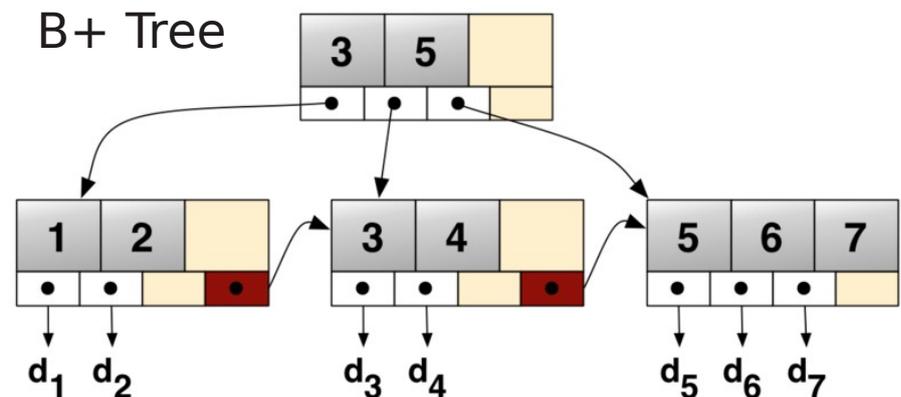
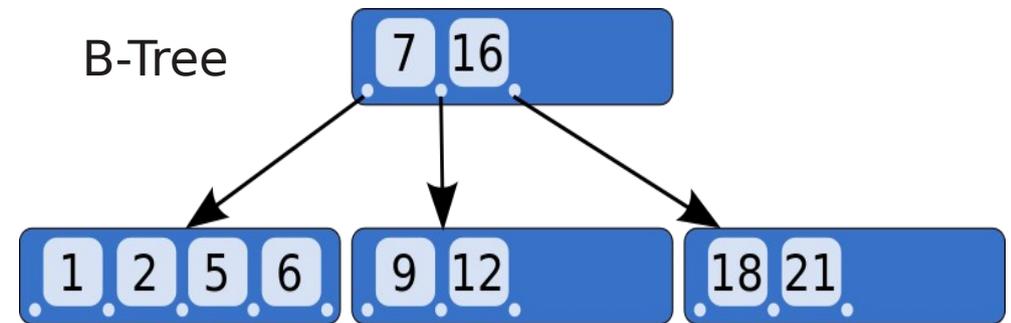
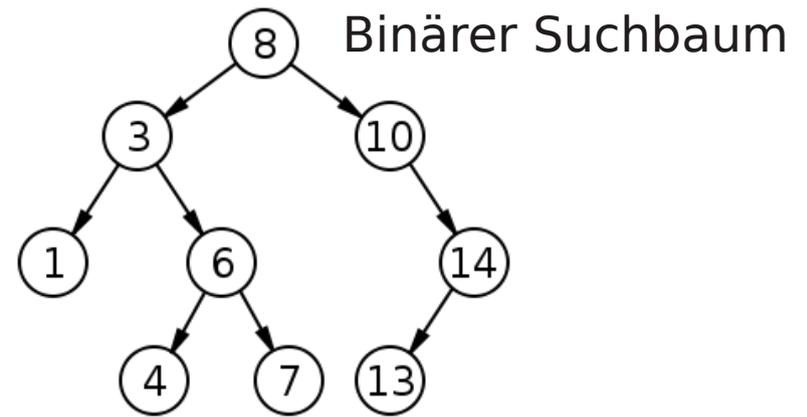
- Instant crash recovery
- Historie
 - „fine-grained“ 30-60 sekunden ohne snapshot
 - „coarse-grained“ bis 60 Tage dank snapshot
- PFSs = „pseudo-file-systems“

Spezifikationen

- 64-bit
- Bis zu 1 Exbibyte Speicherkapazität
- Nutzt B+ Tree für Verzeichnisstruktur
- Unterstützt on-demand data deduplication
- Daten und Metadaten mittels CRC-checks auf Integrität geprüft

B+ Tree

- Kann mehr als 2 Kinder haben
- Immer vollständig balanciert
- Eigentliche Daten nur in Blättern gespeichert
- Verwendet von vielen Datei- und Datenbanksystemen



Bilder von Wikipedia

Instant crash recovery

- Nach crash kein fsck nötig
- Nutzt Historie um letzten stimmigen Zustand wiederherzustellen
- geschieht zur mount-Zeit
 - In weniger als einer Sekunde

Historie - fine-grained

- Genaue Historie wird gespeichert
- Alle 30-60 Sekunden eine prune Operation
 - Prune löscht die Historie bis zum letzten Snapshot
- Dadurch sind entsprechend 30-60 Sekunden feine Historie verfügbar

Historie - coarse-grained

- HAMMER speichert
 - 1 Tag lang 30-sekunden Snapshots
 - 60 Tage lang tägliche Snapshots
- Live Zugriff möglich
- UNDO ermöglicht wiederherstellen einer beliebigen Version
 - Alternativ auch diff output möglich
- Nächtlicher Cronjob umfasst neben Snapshot auch rebalance, dedup und reblock Commands

Data Deduplication

- Doppelt vorkommende Daten werden nur einmal gespeichert
- Dafür zwei Referenzen auf den gleichen Datensatz
- Spart 10% - 40% Speicherplatz
- Optional in HAMMER

Data Deduplication

- Findet auf Block level statt
- Vergleich durch CRC-32
 - Kein zusätzlicher Overhead, da CRC-checks ohnehin für Datenintegrität durchgeführt werden
- Momentan nur per PFS, (noch) nicht übergreifend

PFSs

- „Pseudo File System“
- HAMMER FS kann in bis zu 65536 PFSs aufgeteilt werden
- Teilen sich den verfügbaren Speicherplatz dynamisch
- Master und slave
 - Slave immer read-only, abgesehen von HAMMER eigenen Operationen

Mirroring

- Funktioniert über PFSs
- Master oder Slave wird auf Slave gespiegelt
- Mehrere Slave PFSs als Mirror sind möglich
 - Mehrere PFSs auf einen Mirror zu spiegeln ist nicht möglich

Multi-Volume

- Mehrere Physikalische Volumes pro HAMMER FS
- Bis zu 256 Volumes
- Bis zu 4 Petabytes = 4096 Terabytes pro Volume

Vergleich

- Langsamer als ZFS
 - Aber schneller als UFS
- DragonFlyBSD eigene Entwicklung
 - ZFS ist ein Port von Solaris
- Mehr Features als UFS, weniger als ZFS
- Vergleichsweise jung

Fazit

- Jung
 - Neu
 - Gute Ideen
 - Für sein Alter bereits gute Performance und großes Featureset
- => Interessantes FS in der Welt der BSDs, auf das man ein Auge werfen sollte

Vielen Dank für die Aufmerksamkeit

Quellen

- <http://www.dragonflybsd.org/hammer/>
- <http://leaf.dragonflybsd.org/cgi/web-man?command=hammer§ion=5>
- <http://leaf.dragonflybsd.org/cgi/web-man?command=hammer§ion=8>
- http://leaf.dragonflybsd.org/cgi/web-man?command=newfs_hammer§ion=8
- https://en.wikipedia.org/wiki/DragonFly_BSD
- <https://en.wikipedia.org/wiki/HAMMER>
- https://en.wikipedia.org/wiki/Unix_File_System
- <https://en.wikipedia.org/wiki/ZFS>
- https://en.wikipedia.org/wiki/Binary_Search_Tree
- <https://en.wikipedia.org/wiki/B-tree>
- https://en.wikipedia.org/wiki/B%2B_tree
- http://kerneltrap.org/DragonFlyBSD/HAMMER_Crash_Recovery
- http://www.phoronix.com/scan.php?page=article&item=dragonfly_hammer

Inhaltsverzeichnis

- Allgemein
 - Was ist HAMMER überhaupt?
 - DragonFlyBSD
 - Was kann HAMMER denn so?
- Features genauer betrachtet
 - Spezifikationen
 - B+ Tree
 - Instant crash recovery
 - Historie
 - Data Deduplication
 - PFSs
 - Mirroring
 - Multi-Volume
- Vergleich
- Fazit