

# EXT4

Moderní souborové systémy

Jaroslav Ráb

KIV/PD

# Úvod

- \* EXT4

- \* Fourth extended file system

- \* Žurnalovací souborový systém

# Historie EXT4

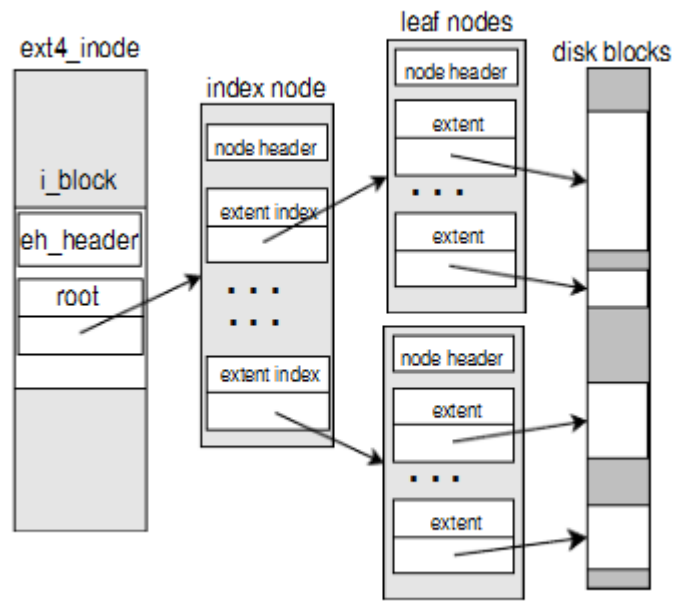
- \* První verze EXT v roce 1992
- \* Začátek vývoje EXT4 v roce 2006 (fork EXT3)
- \* Poprvé podpora v jádře ve verzi 2.6.19 – říjen 2006
- \* Stabilní verze 2.6.28 - prosinec 2008
  
- \* Vývojáři: Mingming Cao, Andreas Dilger, Alex Zhuravlev (Tomas), Dave Kleikamp, Eric Sandeen, Sam Naghshineh, Theodore Ts'o, others

# Vlastnosti EXT4

- \* Max. velikost oddílu: 1 EiB
- \* Max. velikost souboru: 16 TiB pro 4KiB bloky
- \* Max. počet souborů: 4 miliardy
- \* Max. délka názvu souboru: 256 B
- \* Max. počet podadresářů: neomezeno
- \* Není přímá podpora non-linux OS
- \* Obsah adresáře: Spojový seznam, Hash tree
- \* Alokace souborů: Extents/Bitmap
- \* Chybné bloky: Tabulka

# Struktura

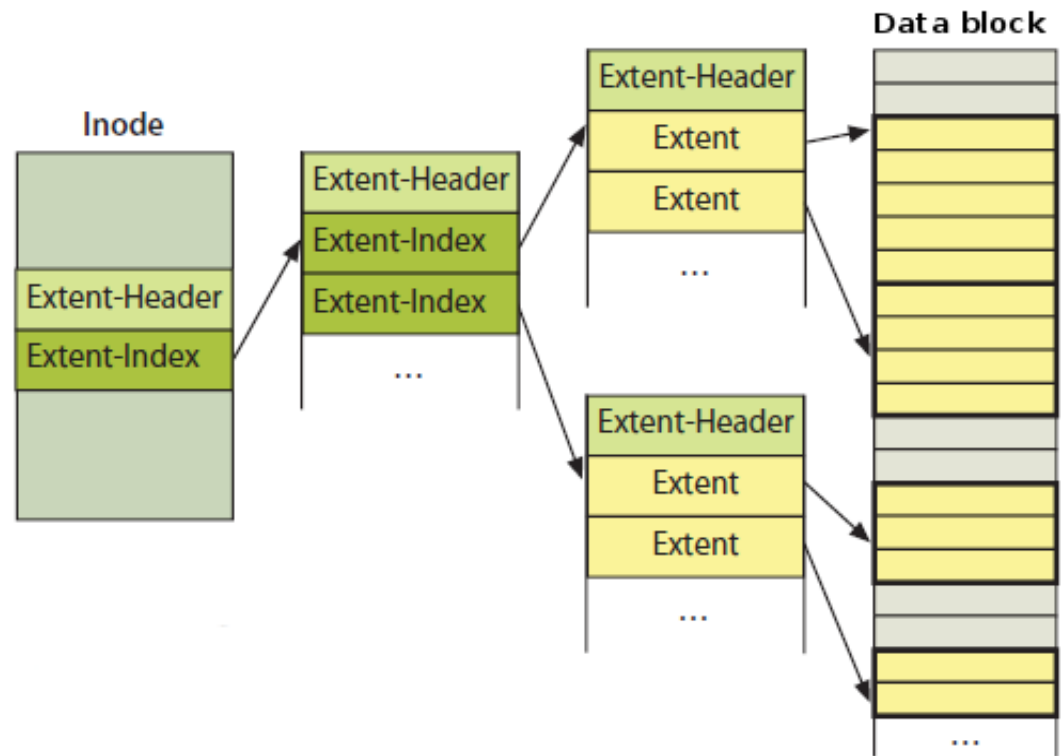
- \* Blok 4KiB
- \* Group 32 768 bloků = 128MiB
- \* Bloky LE
- \* Žurnál BE
- \* Inode 256 B



Group 0 Padding	ext4 Super Block	Group Descriptors	Reserved GDT Blocks	Data Block Bitmap	inode Bitmap	inode Table	Data Blocks
1024 bytes	1 block	many blocks	many blocks	1 block	1 block	many blocks	many more blocks

# Extents

- \* Namapování velkého rozsahu fyzických bloků na disku na jeden ukazatel
- \* Až 128 MiB velký prostor s velikostí bloků 4 KiB



# Persistent pre-allocation

- \* Není potřeba prealokace bloků zápisem nul .
- \* Přidáno systémové volání *preallocate()*.
- \* Místo vzniklé *preallocate()* by nemělo být fragmentované.

# Delayed allocation

- \* Pozdržení alokace bloků do doby než známe přesnou velikost.
- \* Omezení fragmentace souborů.
- \* Nevýhoda při výpadku napájení.



# Multiblock allocator

- \* Namapování velkého množství bloků, učených pro zápis pro jeden soubor = jedna operace.
- \* Nalezení nejvhodnějšího místa pro soubor.
- \* Omezení fragmentace.
- \* Odeslání více bloků najednou.

# Přesnější timestamp

- \* Jádro 2.6.23 přidává nanosekundy ke stávajícímu timestampu v systému. (k sekundové části přidaly 2 bity)
- \* EXT4 s tím již počítá a využívá.
- \* Posunutí problému roku 2038 o dalších 500 let.

# Kontrola souborového systému a Nemožnost obnovení smazaných souborů

- \* Program e2fsck ignoruje bloky nealokovaného místa a tabulku inodů.
- \* Rychlejší kontrola celého souborového systému.
- \* V kořenovém adresáři – Adresář „lost+found“
- \* Není možnost jak obnovit smazané soubory.

# Žurnál a Online defragmentace

## \* Žurnál

- \* žurnál - metadata i obsah souborů. Nejspolehlivější, ale zároveň nejpomalejší.
- \* writeback - metadata se žurnálují, ale obsah souborů ne. Toto je nejrychlejší způsob, ale přináší riziko.
- \* ordered - podobné jako writeback. Nejdříve zapsání souboru na disk, pak v metadatach označí jako zapsaná. Kompromisem mezi výkonem a stabilitou. Výchozí.
- \* vypnuto - větší výkon, nebezpečí

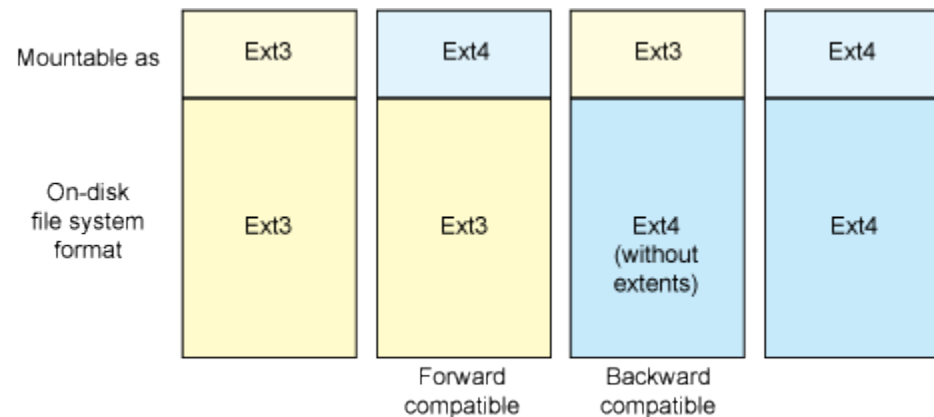
## \* Online defragmentace

- \* I přes podporu Extents a Delayed allocation je potřeba

# Vytvoření EXT4 oddílu

- \* Potřeba minimální verze jádra 2.6.28
- \* Díky podpoře v jádře velice jednoduché
- \* Zpětná a dopřední kompatibilita (vyjma Extents)

- \* Migrace z ext3 na ext4
  - \* S konverzí nebo bez



# Vhodnost pro SSD

- \* Použitelné.
- \* Možnost vypnout žurnál, ale není potřeba.
- \* Defragmentace je ale zbytečná. (nevhodná)

# Reference

- \* V lednu 2010 upgradoval Google své disková uložení z EXT2 na EXT4.
- \* V prosinci 2010 nasazuje Google EXT4 místo YAFFS v mobilním operačním systému Android v 2.3.
- \* V roce 2008 hlavní vývojář EXT3 a EXT4 Theodore Ts'o prohlásil, že ačkoli EXT4 má vylepšené funkce, není to velký pokrok a stále používá staré technologie. A sám věří, že Btrfs je lepší volba, protože nabízí lepší škálovatelnost, spolehlivost.

# Užitečné odkazy

- \* Výkonnostní test EXT3, EXT4, Btrfs, XFS, JFS, ReiserFS, NILFS2 na HDD a SSD:
- \* [http://www.phoronix.com/scan.php?page=article&item=linux\\_2638\\_large&num=1](http://www.phoronix.com/scan.php?page=article&item=linux_2638_large&num=1)
- \* Návod pro nasazení:
- \* [https://ext4.wiki.kernel.org/articles/e/x/t/Ext4\\_Howto\\_doob.html](https://ext4.wiki.kernel.org/articles/e/x/t/Ext4_Howto_doob.html)



# Zdroje

- \* <http://www.root.cz/clanky/ext4-evolucni-souborovy-system/>
- \* <https://wiki.archlinux.org/index.php/Ext4>
- \* <http://cs.wikipedia.org/wiki/Ext4>
- \* <https://ext4.wiki.kernel.org/index.html>
- \* <http://marc.info/?l=linux-ext4&m=125803982214652&w=2>
- \* <http://www.h-online.com/open/features/The-Ext4-Linux-file-system-746579.html>
- \* <http://www.kernel.org/doc/Documentation/filesystems/ext4.txt>
- \* <http://arstechnica.com/open-source/news/2009/04/linux-collaboration-summit-the-kernel-panel.ars>
- \* <http://www.slideshare.net/Clogeny/ext3ext4-file-systems>
- \* <http://atrey.karlin.mff.cuni.cz/~jack/papers/lk2009-ext4-btrfs.pdf>
- \* <http://www.scribd.com/doc/7114603/Ext4-File-System>

# Děkuji za pozornost

\* Otázky

Jaroslav Ráb