System plików ext4

Bartosz Borkowski

Systemy Rozproszone

29 kwietnia 2010

Agenda

- Historia
- Pierwsze kroki
- ext4 od środka
- Testy
- Podsumowanie

Historia

- 1987r. Minix V1 FS
- 1992r. Extended file system
- 1993r. ext2
- 1999r. ext3
- 2001r. ReiserFS
- 2006r. ext4

Pierwsze kroki

Z wiedzy tajemnej Linusa

ext3 _sucks_ in many ways. It has huge inodes that take up way too much space in memory. It has absolutely disgusting code to handle directory reading and writing (buffer heads! In 2006!). It's conditional indexing code is horrible. Its performance absolutely sucks when the journal is being drained or something.

Pierwsze kroki

Z wiedzy tajemnej Linusa

ext3 _sucks_ in many ways. It has huge inodes that take up way too much space in memory. It has absolutely disgusting code to handle directory reading and writing (buffer heads! In 2006!). It's conditional indexing code is horrible. Its performance absolutely sucks when the journal is being drained or something.

- W 2006r. rozpoczęto prace nad ext4, biorąc za bazę kod ext3,
- zmiany po raz pierwszy włączono do jądra 2.6.19,
- 11 października 2008 ext4 uznano za stabilny i włączono do jądra 2.6.28.

Rozmiar

- teoretycznie obsługuje woluminy do 1 EiB (2^{60} bajtów),
- teoretycznie obsługuje pliki do 16 TiB,
- praktycznie system plików nadal jest ograniczony do 16 TiB,
- oferuje dwukrotnie więcej podkatalogów.

Extent

- extent zastępuje bloki z poprzednich wersji systemu,
- jeden extent mapuje do 128MB,
- współdziała z mechanizmem H-drzew.

Alokacja

Pre-alokacja

Nowa funkcja systemowa gwarantująca najprawdopodobniej ciągły obszar dla pliku.

Opóźniona alokacja

Technika optymalizacyjna polegająca na opóźnianiu zrzutu danych na dysk aż stanie się to niezbędne.

Wieloblokowa alokacja

Naturalne rozszerzenie opóźnionej alokacji: jednoczesne alokowanie wielu bloków podczas zrzutu danych.

Księgowanie

Sumy kontrolne

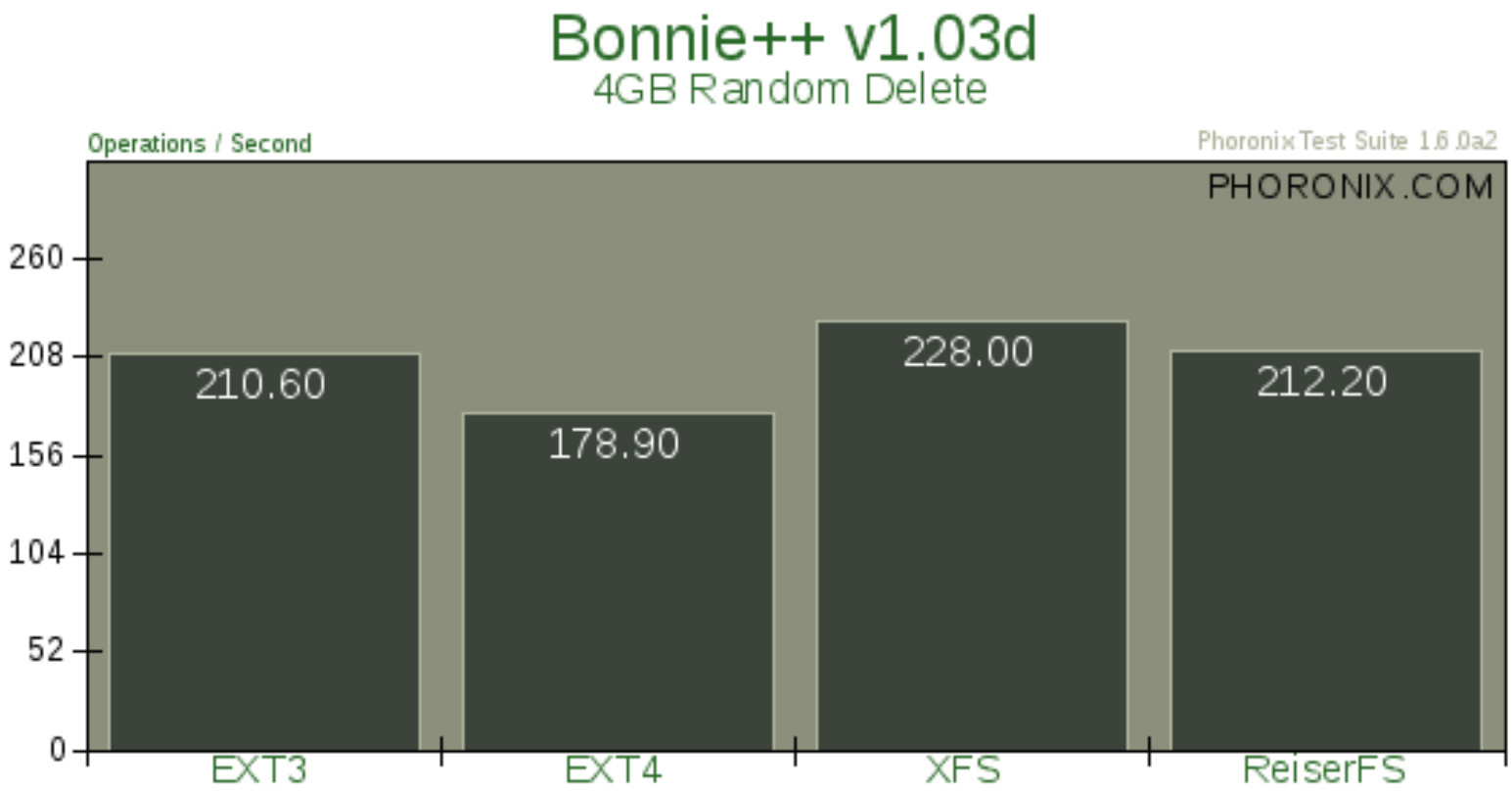
- każda transakcja
- każdy deskryptor grupy bloków

Bariery

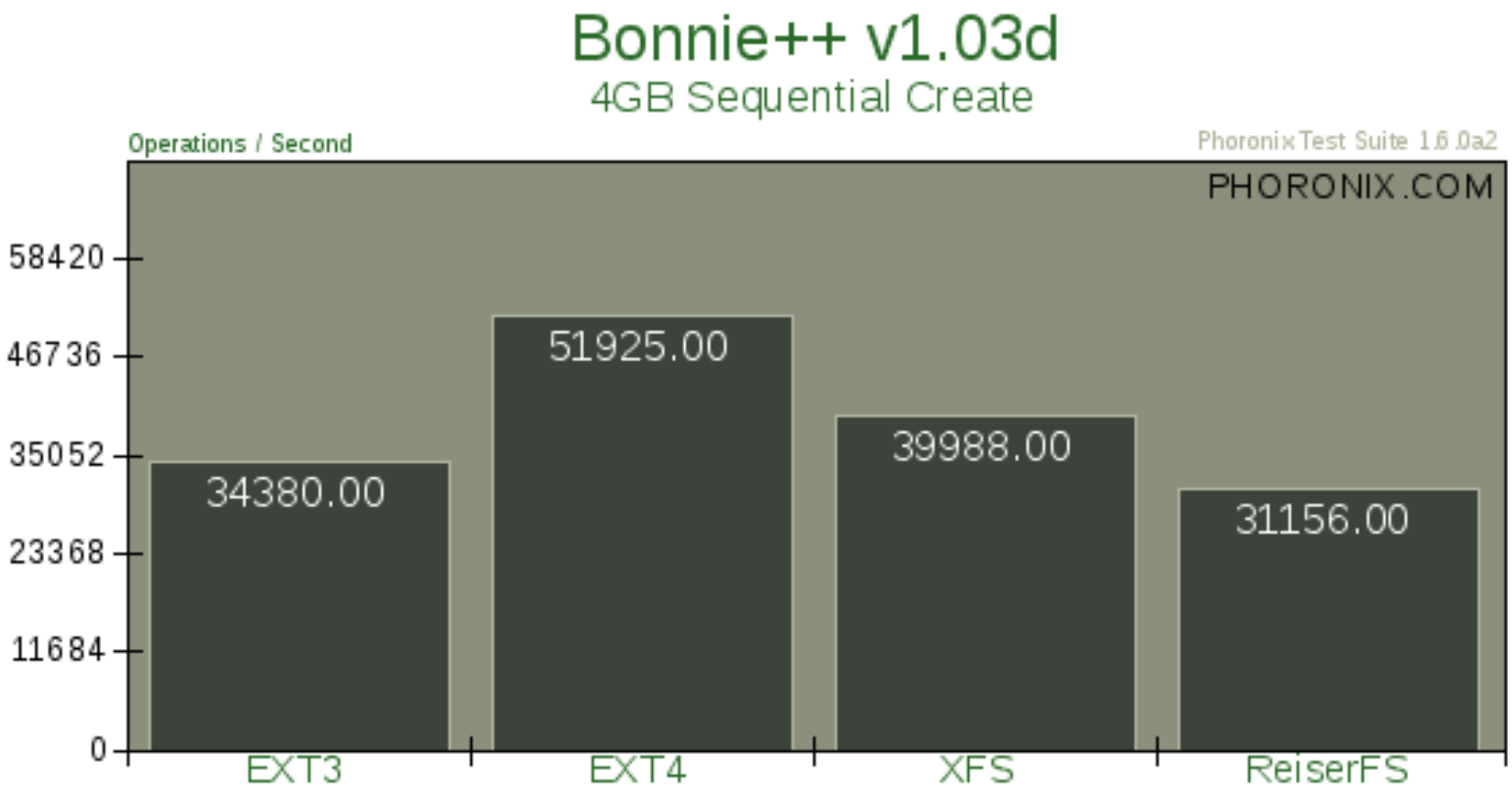
Wprowadzono mechanizm barierowy, który nakazując sterownikowi zapis danych w określonym porządku, zapobiega rozspójnieniu danych.

Timestamps

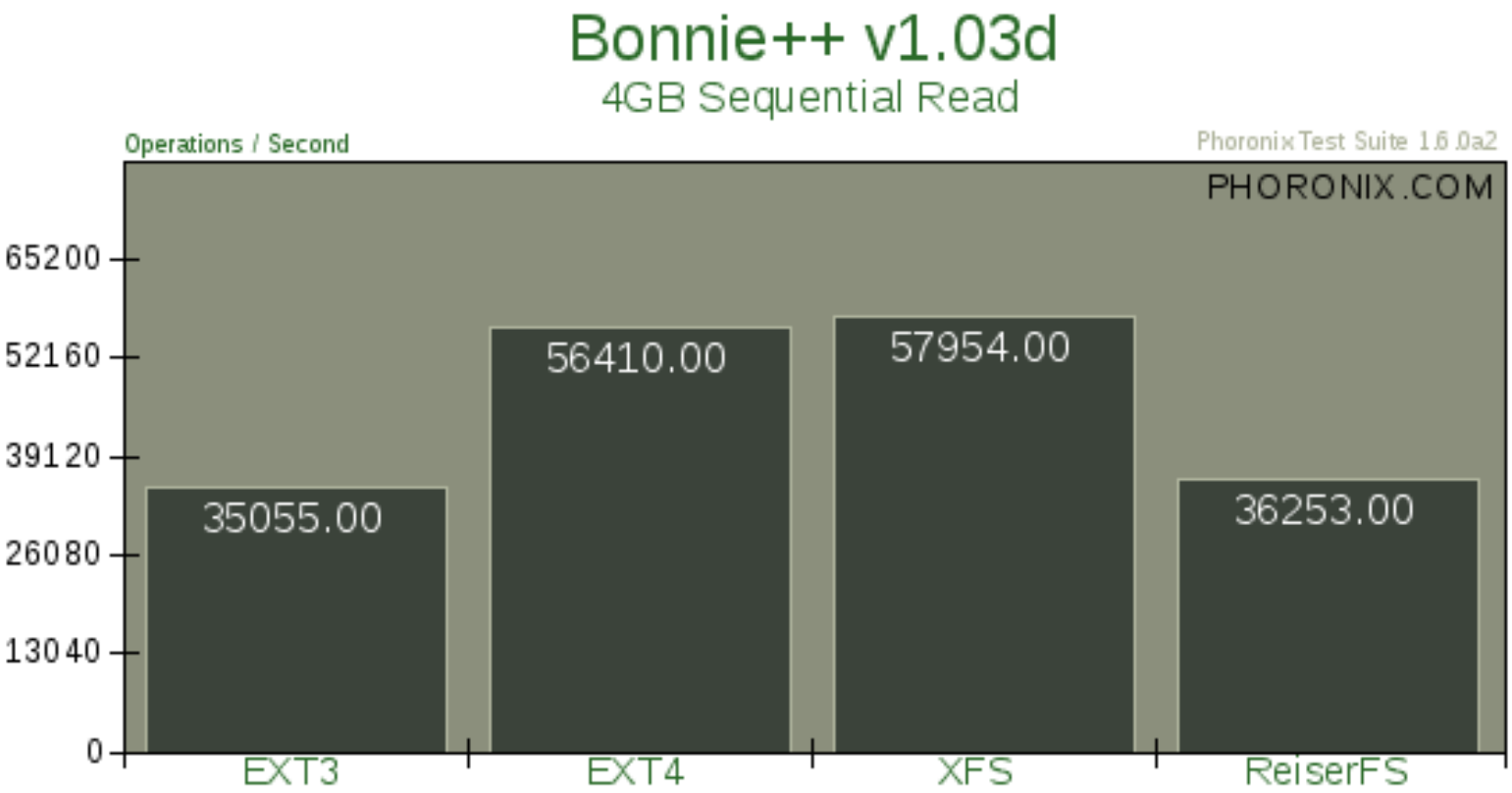
- większa rozdzielczość,
- problem roku 2038,
- nowy timestamp.



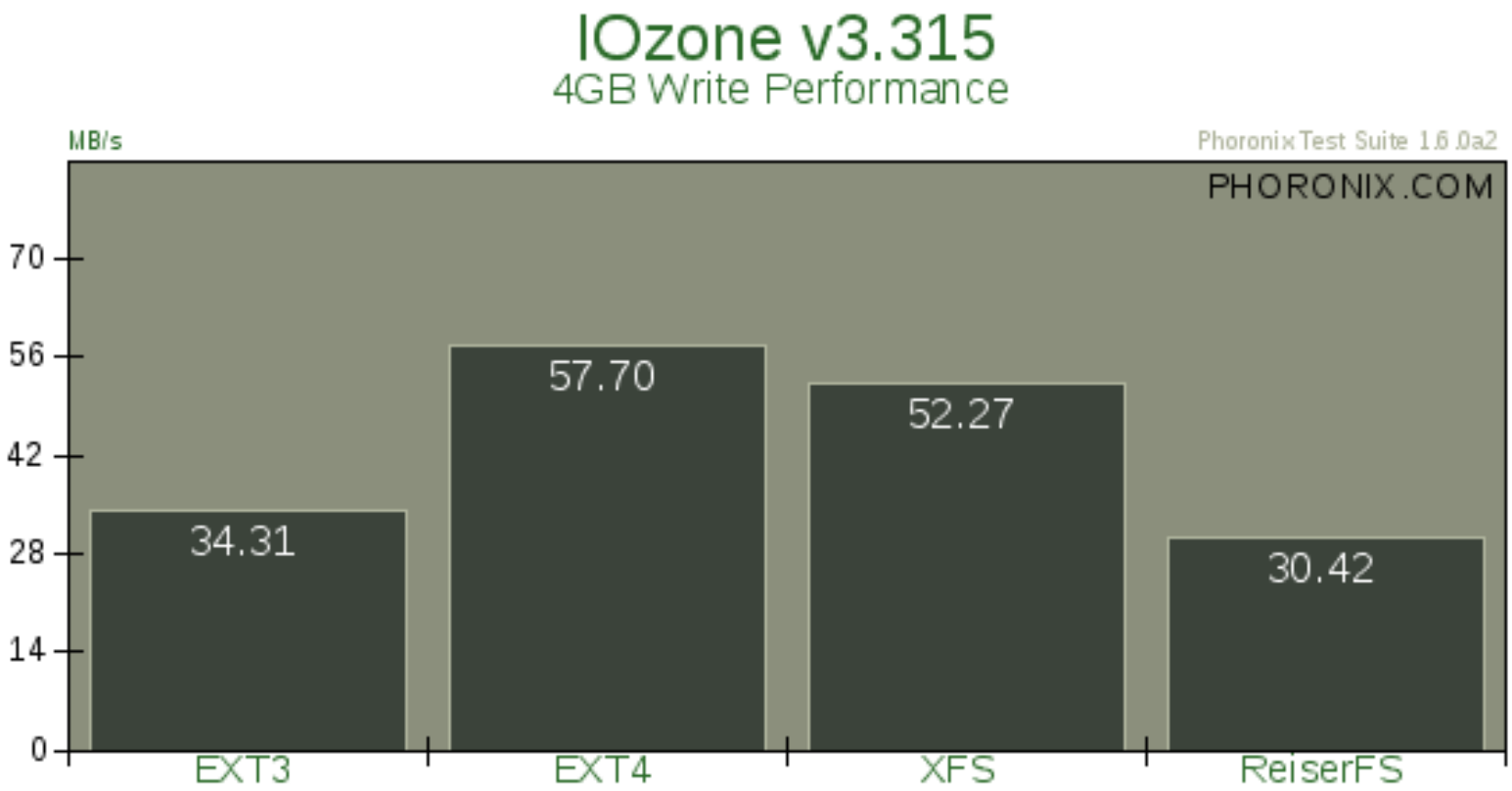
http://www.phoronix.com/scan.php?page=article&item=ext4_benchmarks



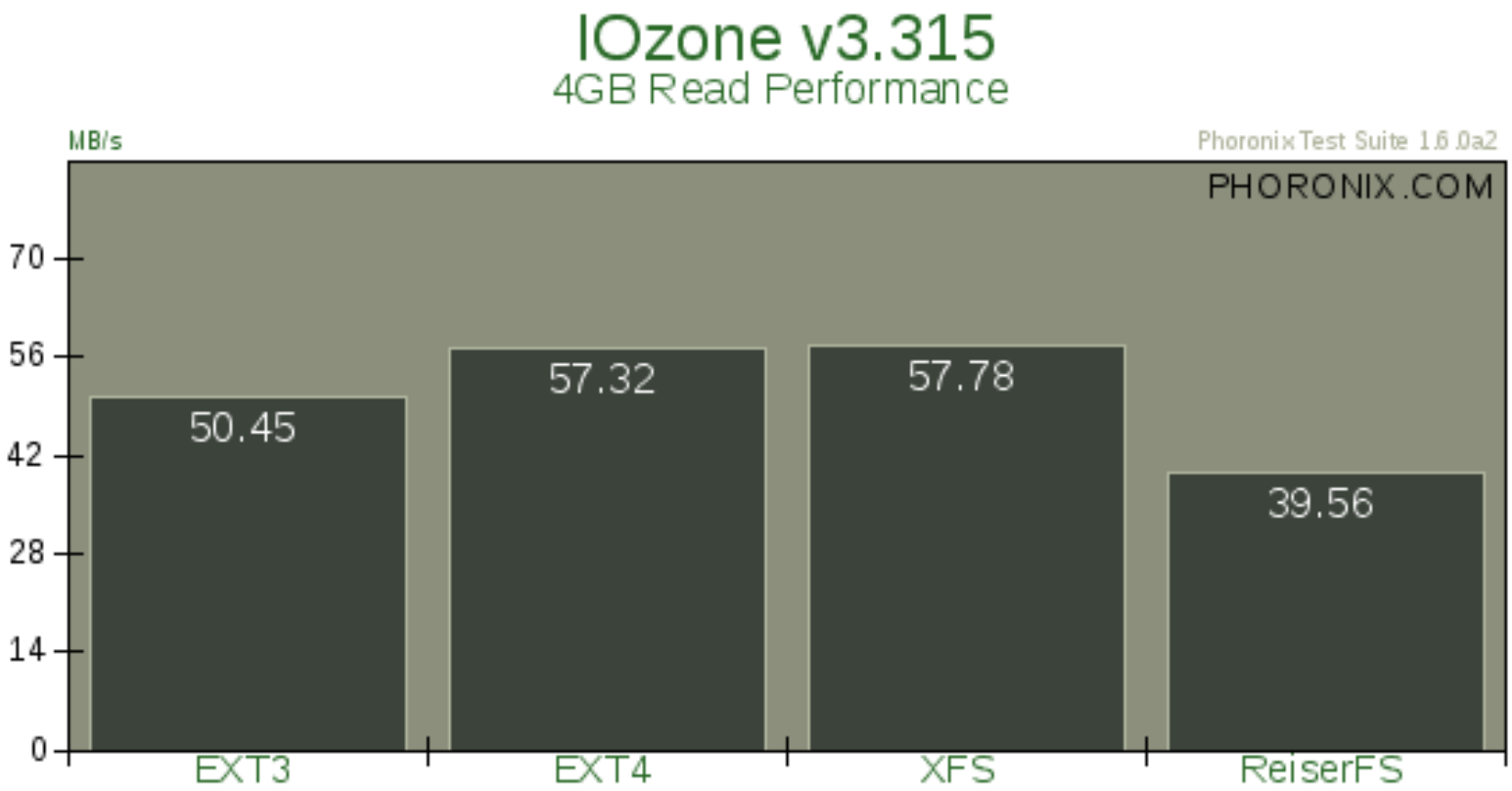
http://www.phoronix.com/scan.php?page=article&item=ext4_benchmarks



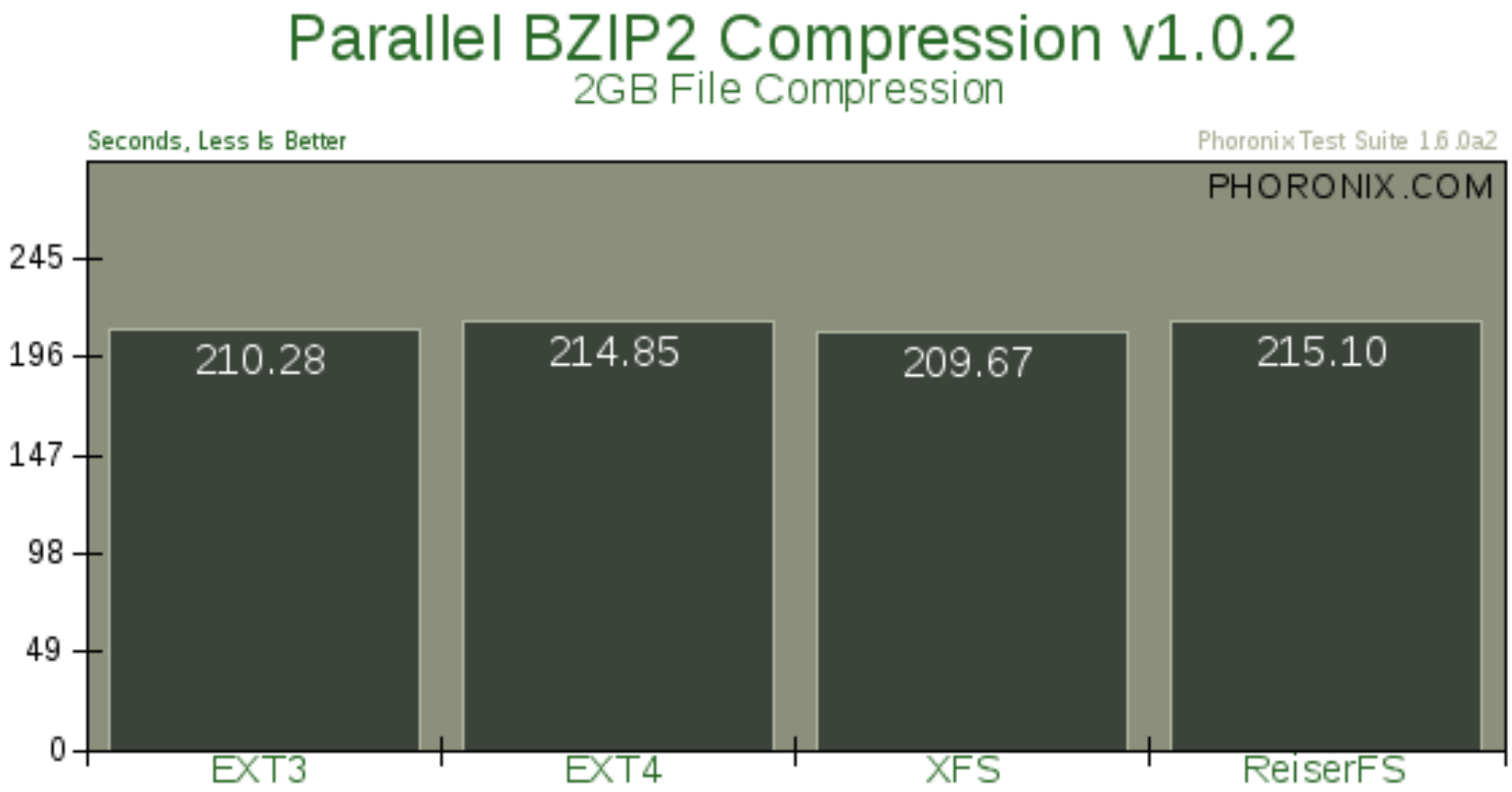
http://www.phoronix.com/scan.php?page=article&item=ext4_benchmarks



http://www.phoronix.com/scan.php?page=article&item=ext4_benchmarks



http://www.phoronix.com/scan.php?page=article&item=ext4_benchmarks



http://www.phoronix.com/scan.php?page=article&item=ext4_benchmarks

Co dalej?

- Google już ogłosiło, że będzie używać ext4,
- kilkanaście znanych bugów,
- spora lista TODO.

- https://ext4.wiki.kernel.org/index.php/Main_Page
- http://lkml.org/
- http://kernelnewbies.org/
- http://www.phoronix.com/