

Ext3

Wprowadzenie

Kompatybilność

System Ext3 jest kompatybilny w przód i w tył z Ext2, czyli

- System plików Ext2 można zamontować jako Ext3
- System plików Ext3 można zamontować jako Ext2 o ile został prawidłowo odmontowany

Podstawowe cechy

System Ext3 - ze względu na zgodność z Ext2 - ma takie same właściwości i ograniczenia jak jego poprzednik.

Cecha	Wartość dla Ext3
Maksymalna długość nazwy pliku	255 bajtów
Znaki, które mogą wystąpić w nazwie pliku	Wszystkie z wyjątkiem 0x00
Maksymalna długość ścieżki	Brak
Maksymalna długość pliku	2 TB (2^{41} bajtów)
Maksymalny rozmiar partycji	32 TB (2^{45} bajtów)
Rozmiar bloku	1024, 2048 lub 4096 bajtów
Inne cechy wspólne z Ext2	<ul style="list-style-type: none">• Obsługa dowiązań symbolicznych i twardych• Rozróżnianie wielkości liter• POSIXowy model uprawnień• Zapisywanie czasu utworzenia, dostępu i modyfikacji pliku• Obsługa kompresji za pomocą odpowiednich nakładek

Różnice względem Ext2

- Kronikowanie (*journaling*)
- Indeksowanie katalogów (*Hashed tree*)
- Fragmentacja bloków
- Zerowanie wskaźników do bloków przy kasowaniu
- Opóźnienie zapisywania danych

Kronikowanie

System ext3 nie posiada wbudowanego mechanizmu kronikowania. Zamiast tego korzysta z JBD (*Journaling block device*), który w pewnym sensie stanowi API dzięki któremu do dowolnego systemu plików możemy dodać kronikowanie i transakcyjność.

W ext3 wykorzystywany jest dziennik o działaniu podobnym do dziennika „redo” w bazach danych - to znaczy zapisywane są w nim tylko nowe wartości do zapisania na dysku. W przypadku awarii systemu operacje z dziennika wykonywane są ponownie.

System ext3 posiada następujące trzy tryby kronikowania:

- *Writeback*:

Najszybszy tryb. Do dziennika trafiają jedynie modyfikowane metadane, w związku z tym nie jest zapewnione bezpieczeństwo danych, a jedynie spójność systemu plików. Mamy gwarancję, że w przypadku awarii systemu czas naprawy systemu plików nie będzie zależał od rozmiaru partycji.

- *Ordered*:

Domyślnie stosowany tryb. Podobnie jak *writeback*, kronikowane są jedynie metadane, jednakże zanim zostaną zapisane we właściwe miejsce na dysku, najpierw zapisywane są właściwe dane, których dotyczą. Jest nieco wolniejszy niż poprzedni tryb, ale zabezpiecza przed sytuacją, w której przy dopisywaniu danych na końcu pliku zostają zapisane śmieci (ma to miejsce, jeśli awaria nastąpiła po zapisaniu metadanych, ale przed zapisaniem odpowiednich bloków z danymi).

- *Journal*:

Tryb z pełnym kronikowaniem - zarówno danych jak i metadanych. Gwarantuje największe bezpieczeństwo, ale i najmniejszą wydajność.

Domyślnie dziennik jest przechowywany w katalogu głównym w pliku *.journal*, ale teoretycznie można go umieścić w dowolnym innym miejscu - na przykład na innym, szybszym dysku.

Oprócz dziennika, w superbloku w ext3 trzymana jest lista osieroconych (*orphaned*) plików. Znajdują się na niej dane o plikach, które należy skasować przy montowaniu systemu. Taka lista jest konieczna, gdyż w momencie awarii systemu mogą być otwarte, ale skasowane pliki - normalnie powinny one zostać skasowane zaraz po ich zamknięciu - dlatego w przypadku awarii ta informacja musi być gdzieś pamiętana.

Indeksowanie katalogów

Dla katalogów z bardzo dużą ilością plików można włączyć alternatywny sposób przechowywania ich zawartości. Polega on na zapisywaniu pozycji w katalogu w drzewie indeksowanym wartościami funkcji hashującej z nazwy wpisu. Umożliwia to szybko (w czasie logarytmicznym) odnajdowanie pliku bez konieczności ładowania i przeglądania całej zawartości katalogu.

Fragmentacja bloków

System ext3 pozwala na podział bloku na kawałki równej wielkości, w których można przechowywać różne pliki. Znacznie zmniejsza to fragmentację wewnętrzną w przypadku dużych ilości małych plików.

Zerowanie wskaźników

Przy kasowaniu pliku zerowane są wskaźniki do węzłów, przez co w ext3 praktycznie niemożliwe jest odtworzenie skasowanych plików.

Opóźnienie zapisu danych

Zapisy na dysk odbywają się co 5 sekund zamiast co 30 jak w Ext2

Ext3 w innych systemach operacyjnych

Do systemu plików zapisanego w formacie ext3 można dostać się również z systemów, które nie mają wbudowanej obsługi dla niego. Niestety, w większości przypadków jest to dostęp tylko do odczytu. Jednym z wyjątków jest program Explore2fs dla Windows 9X / NT, który pozwala również na zapis. Dostępne są także nakładki na systemy operacyjne dzięki którym ext3 jest widziany jako normalna partycja dla wszystkich aplikacji.

Podsumowanie

Odporność na awarie	Wysoka - możliwość kronikowania zarówno danych jak i metadanych
Transakcyjność	Tak - operacje na plikach są atomowe
Obsługa dużych dysków i dużych plików	Praktycznie brak ograniczeń
Prawa dostępu i inne metadane	Rozbudowane: właściciel, uprawnienia, data i czas dostępu, modyfikacji i utworzenia, alternatywne strumienie
Długie nazwy plików i ścieżek	Pełne wsparcie
Linki	Zarówno symboliczne jak i twarde
Organizacja plików	Głównie listy i bitmapy. Możliwość użycia H-Drzew
Kompresja	Tylko po dodaniu odpowiedniej łątki
Fragmentacja	Prealokacja
Szyfrowanie	Brak
Wersjonowanie	Brak

Źródła

- [Whitepaper: Red Hat's New Journaling File System: ext3](#)
- [EXT3, Journaling Filesystem](#)
- [Advanced filesystem implementor's guide, Part 1](#)
- [Advanced filesystem implementor's guide, Part 7](#)
- [Advanced filesystem implementor's guide, Part 8](#)
- [Wikipedia: Comparison of file systems](#)