

## 1. Powrót do przeszłości czyli EXT2

Drugi Rozszerzony System Plików ( Second Extended File System , w skrócie Ext2 ) pojawił się w pierwszej wersji alfa w styczniu 1993r. Kilka lat temu Ext2 był podstawowym systemem plików dla wielu dystrybucji Linuxa. System EXT3 wiele odziedziczył z EXT2, warto więc zapoznać się z zaletami i wadami EXT2.

Zalety:

- Wspiera wszystkie elementy systemu plików Unix (dowiązania symboliczne, pliki specjalne, prawa dostępu...)
- Wysoka wydajność dzięki przeciwdziałaniu fragmentacji (poprzez przydzielanie bliskich bloków oraz prealokację).
- Wydajny mechanizm dowiązań symbolicznych
- Stabilny i dobrze przetestowany (sam system plików, jak również program naprawiający e2fsck)
- Dobrze zdefiniowany sposób dodawania rozszerzeń
- Niezależny od tworzącego systemu operacyjnego (wszystkie pola wielobajtowe zapisane w standardzie Little - endian)
- Maksymalny rozmiar partycji to 4TB, a pojedynczego pliku 2GB. Maksymalna długość nazwy pliku: 255 znaków.
- Obsługa "dziurawych" plików (nieużywane bloki nie zostają przydzielone).
- Automatyczne sprawdzanie systemu po awarii

Wady:

- Mało efektywna obsługa katalogów (choć użycie pamięci podręcznej znacznie ją poprawia)
- Niska wydajność dla bardzo małych plików (rzędu kilkuset bajtów) - duże straty na alokację i stosunkowo wolny dostęp
- Długotrwałe sprawdzanie systemu plików po niepoprawnym zamknięciu systemu W systemie Ext2 jest to automatycznie uruchamiany program fsck, który przy dużej powierzchni dysków może działać długo.

## 2. EXT3 – następca EXT2

Trzeci Rozszerzony System Plików wywodzi się bezpośrednio z Ext2. Został stworzony przez Stephen'a Tweedie z firmy RedHat. Wersja alfa pojawiła się w 1999 roku. Podstawową różnicą w stosunku do poprzednika jest mechanizm księgowania (journaling), który wprowadzono w celu skrócenia czasu wstawiania systemu po awarii. Dodatkowo, wprowadzono na stałe kilka usprawnień, które były wcześniej dostępne jako łaty dla systemu Ext2.

Spośród dodatkowych cech w stosunku do poprzedniej wersji warto wymienić następujące możliwości systemu Ext3:

- Journaling - mechanizm księgowania zwiększa bezpieczeństwo systemu.

- Indeksowane katalogi - znacznie zwiększają wydajność systemu przy dużej ilości plików.
- Zapis synchroniczny - w najnowszych wersjach systemu Ext3 działa ponad 10 razy szybciej od wersji z Ext2 (dotyczy zapisu do zwykłych plików).
- Możliwość przechowywania rozszerzonych atrybutów i list kontroli dostępu POSIX
- Zerowanie wskaźników-utrudnia odnalezienie usuniętych plików
- Fragmentacja bloków- zmniejsza ilość zajmowanego miejsca przez dane

W dalszej części prezentacji znajduje się szczegółowy opis powyżej wymienionych nowości w EXT3

### 3. Kronikowanie

Kronikowanie to taka funkcja systemu plików, która zapisuje operacje zlecone systemowi plików, ale jeszcze nie dokończone w specjalnym pliku nazywanym dziennikiem (ang. journal). Kronikowanie jest pomysłem na złagodzenie skutków błędów spowodowanych przez załamanie się systemu w trakcie wykonywania operacji zapisu. Do tej pory takie błędy powodowały konieczność uruchomienia specjalnego programu, który przywracał spójność danych. Było to jednak bardzo powolne, gdyż wymagało przejrzenia całego dysku. Ideą kronikowania jest zastąpienie tego przez możliwość sprawdzenia jedynie ostatnich operacji zapisu odnotowanych w specjalnym dzienniku.

System plików z kroniką po restarcie zwykle przywraca spójność dysku w czasie rzędu kilku, kilkunastu sekund.

Restart systemu po awarii w systemach z kronikowaniem nie zależy od wielkości partycji, co jest szczególnie istotne w przypadku wielkich dysków.

Kronikowanie nie zapewnia jednak, że pewne dane nie zostaną utracone.

#### Scenariusz zapisu danych

Zazwyczaj scenariusz zapisu danych (lub metadanych) na dysku wygląda następująco:

- umieszczamy wpis w dzienniku co zmienimy i w jaki sposób tego dokonamy (aby nie zakończone zmiany można było wycofać i przywrócić stan po awarii).
- zapisujemy zmiany na dysku
- opcjonalnie usuwamy wpis z dziennika

#### Odtwarzanie spójności systemu plików

Podczas odtwarzania spójności systemu plików program e2fsck rozróżnia dwa przypadki:

- Załamanie systemu nastąpiło przed zakończeniem zapisywania operacji do dziennika - w dzienniku jest niekompletny wpis, ale struktury systemu plików pozostają niezmienione, wtedy e2fsck zignoruje wpis
- Załamanie systemu nastąpiło po wpisaniu operacji do dziennika - zapisane bloki są kompletne i e2fsck zapisuje je na dysk, a tym samym przywraca spójność danych

#### Typy kronikowania

Rozróżniamy dwa podstawowe typy kronikowania:

1. tylko metadanych (zapewnia spójność metadanych, ale może nie pamiętać ostatnio wpisanych danych tuż przed awarią)

W systemach plików Ext2 i Ext3 wyróżniamy sześć rodzajów metadanych:

- superbloki
  - deskryptory grup
  - i-węzły
  - bloki do adresowania pośredniego
  - bitmapy bloków danych
  - bitmapy i-węzłów
2. metadanych i danych (odtworzy wszystko co zostało w całości wprowadzone do dziennika)

Większość systemów plików z kronikowaniem, takich jak: ReiserFS, XFS i JFS ograniczają się do zapisywania w dzienniku informacji o zmianach jedynie w metadanych. Jest to wystarczające do odzyskania spójności systemu plików, ale nie zapobiega utracie informacji.

### Trzy tryby kronikowania w Ext3

System plików Ext3 można skonfigurować tak aby zapamiętywał zarówno metadane jak i zwykle dane. Ext3 pozwala administratorowi na wybór spośród trzech trybów kronikowania:

1. **Journal - wszystkie dane oraz metadane** zostaną zapisane do dziennika. Zmniejsza to szanse na utratę danych, jednak zwiększa liczbę koniecznych odwołań do dysku. Jest to jednocześnie **najbezpieczniejsza i najwolniejsza** opcja.
2. **Ordered** - tylko bloki metadanych są kopiowane do dziennika, ale system plików grupuje metadane i odpowiadające im bloki danych tak, aby dane zostały zapisane na dysk przed metadanymi. Zmniejsza to szanse na błędy wewnątrz plików. Jest to domyślny tryb dla Ext3
3. **Writeback** - tylko bloki metadanych są zapisywane do dziennika, jest to najczęściej spotykana opcja w systemach plików z kronikowaniem, jest przy tym najszybsza

### JBD (Journaling Block Device)

System plików Ext3 nie zarządza dziennikiem we własnym zakresie, ale korzysta z wbudowanego w jądro interfejsu do obsługi dzienników zwanego *Journaling Block Device*. Należy pamiętać, że JBD korzysta z tego samego urządzenia co Ext3, zatem jest tak samo podatny na załamania systemu i musi się przed nimi zabezpieczyć.

Współpraca między systemem plików Ext3, a JBD opiera się na trzech podstawowych jednostkach:

1. **Rekordy dziennika (Log record)** - opisują pojedynczy wpis do dziennika (pojedynczą zmianę w systemie plików)
2. **Atomowe operacja (Atomic operation)** - składają się z rekordów dziennika dotyczących tego samego wywołania systemowego
3. **Transakcje (Transaction)** - zawierają kilka atomowych operacji, których rekordy są zaznaczone jako poprawne dla programu e2fsck w tym samym czasie

## **4. Inne nowości w EXT3**

### **Indeksowanie katalogów**

Kolejną nowością, która się pojawiła w EXT3 jest indeksowanie katalogów. W EXT2 wpisy w katalogach były przechowywane na liście liniowej. Było to bardzo nieefektywne, gdy w katalogach znajdowało się bardzo dużo plików. W EXT3 katalogi mają postać drzewa. Operacja wyszukiwania pozycji w katalogu opiera się na obliczeniu funkcji haszującej i zejściu w dół po wskaźnikach w drzewie aż do bloku zawierającego wpis pozycji. Dzięki niej jest możliwe odnalezienie pliku w czasie logarytmicznym (a nie liniowym, jak to było w EXT2). Przy dodawaniu nowej pozycji (przepełnienie bloku z wpisami) stosuje się sortowanie wpisów i podział na dwa bloki. Następnie poprawiane są wpisy w indeksach.

### **Zerowanie wskaźników**

W przeciwieństwie do Ext2, Ext3 zeruje wskaźniki do węzłów usuniętych plików. Konsekwencją tego jest to, iż bardzo trudno odnaleźć usunięte pliki.

### **Fragmentacja bloków.**

System ext3 może podzielić jeden blok na kawałki o tym samym rozmiarze i w każdym z nich przechowywać małe pliki. Pozwala to znacznie zmniejszyć obszar zajmowany przez dane, gdy w systemie jest wiele małych plików.