Souborový systém NTFS (New Technology File System)

Jan Šváb

Historie

- vyvinut Microsoftem pro Windows NT
- postupný vývoj

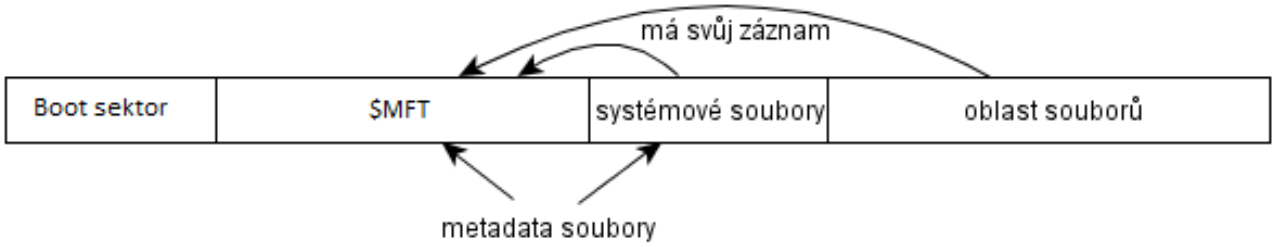
Year	Name	Windows Version	Feature	Max FS Size*
1993	NTFS 1.0	Windows NT 3.1	Journaling	2 TB
1994	NTFS 1.1	Windows NT 3.5		2 TB
1996	NTFS 1.2 jako NTFS 4.0	Windows NT 3.51	Compression, Streams, ACLs	2 TB
2000	NTFS 3.0 jako NTFS 5.0	Windows 2000	Quotas, Encryption, Sparse files, Reparse, DLT	2 TB
2003	NTFS 3.1 jako NTFS 5.1	Windows Server 2003	Expanded MFT, better encryption, Shadow Copies	256 TB
2005	jako NTFS 5.2	Windows Vista	Transactional NTFS, symbolic links	256 TB
2006	jako NTFS 6.0	Windows Server 2008	SMART reader, self healing	256 TB
2009	jako NTFS 6.1	Windows Server 2008 R2	SSD TRIM, native VHD	256 TB

Základní struktura oddílu

- prostor v oddíle rozdělen na **clustery**
  - nejmenší adresovatelné místo z pohledu NTFS
  - skládá se z 2, 4, 8, 16 nebo 32 sektorů
  - max. velikost 64KB
    - fragmentace oddílu vs nevyužité místo v clusteru
- až 64-bitové adresování clusterů
  - teoreticky možné adresovat až 16EB při clusteru 64KB
- logického číslování clusterů (LCN)
  - clustery očíslovány od 0 sekvenčně od začátku oddílu
  - FA (offset v oddílu) = ČÍSLO_C * VEL_C
- clustery souboru identifikovány virtuálním číslováním (VCN)
  - číslovány sekvenčně od 0, ale ve skutečnosti musí jít za sebou
  - součástí souboru mapa VCN->LCN

Základní struktura oddílu

- organizace oddílu



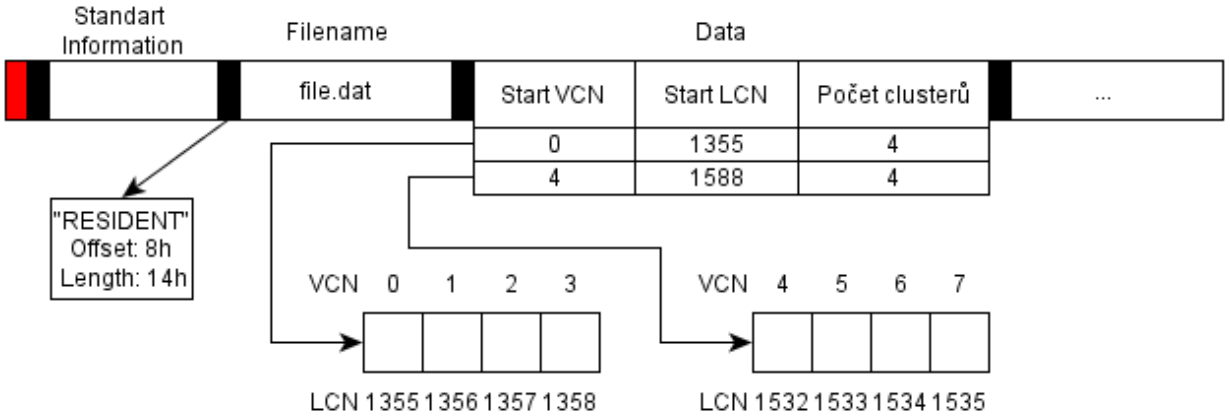
Master File Table ($MFT)

- vytvořen během formátování
- každý soubor(adresář) zde svůj záznam
  - velikost záznamu 1KB obvykle
  - malým souborům (do 512B) stačí => rychlý přístup
- prvních 16 záznamů určeno pro metadata soubory popisující oddíl
  - názvy začínají znakem $
  - skryté (zobrazení např. příkazem dir \ah <NAZEV>)
- záznam souboru identifikován 64bitovým číslem **File Reference**
  - první část = pozice hlavního záznamu v $MFT
  - druhá část = sekvenční číslo pro kontrolu konzistence, inkrementováno vždy po znovupoužití záznamu na dané pozici

Přehled metadata souborů

0	$Mft	Odkaz na sebe (adresa v Boot sektoru)
1	$MftMirr	Záloha prvních 16 záznamů k obnově porušené $Mft. Umístěn uprostřed oddílu nebo na konci v závilosti na OS.
2	$LogFile	Žurnál, zaznamenány poslední provedené transakce pro obnovu
3	$Volume	Informace o oddílu např.: jméno, verze NTFS,...
4	$AttrDef	Definice atributů souborů a jejich popis
5	.	Index souborů a podadresářů v kořenovém adresáři
6	$Bitmap	Bitmapa s informacemi o volných clustrech (1 bit = 1 cluster)
7	$Boot	Ukazatel na nebo kopie kódu pro zavedení OS v Boot sektoru
8	$BadClus	Soubor, který obsahuje vadné cluster. NTFS detekuje vadné clustery, realokace dat a zahrnutí vadného clusteru do $BadClus
9	$Secure	Unikátní bezpečností deskriptory pro soubory
10	$Upcase	Tabulka pro konverzi malých písmén názvů na velká písmena v Unicode kódování
11	$Extend	Adresář pro různá rozšíření NTFS (např.: $Quota)
12-15		Pro budoucí rozšíření

$MFT - záznam



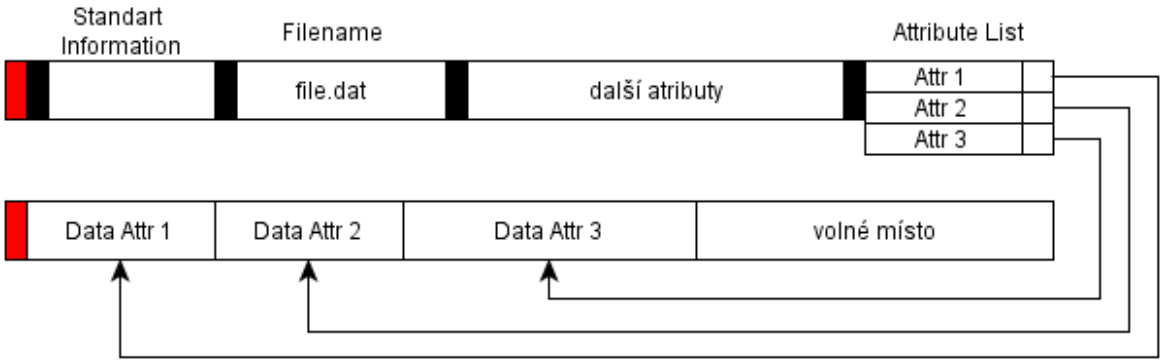
- záznam tvořen hlavičkou a seznamem atributů
- soubor = množina atributů, každý jako samostatný stream bytů
- atribut – pár identifikátor atributu/hodnota
  - popisuje data nebo vlastnosti souboru
  - rezidentní
    - celý součástí MFT záznamu (jméno souboru, timestamp,...), některé vždy
  - nerezidentní
    - hodnota nevejde do záznamu
    - místo hodnoty seznam ukazatelů na tzv. run nebo extent oblasti
    - mapování pomocí VCN

$MFT - záznam

- atribut má volitelné jméno (např.: $FILENAME)
  - v záznamu reprezentován číselným kódem s referencí na $AttrDef

Typ atributu	**Popis**
Standard Information	Časové razítko, počet odkazů,...
Filename	Jméno souboru
Data	Data souboru
Attribute List	Seznam všech dodatečných záznamů s atributy, které se nevešly do hlavního záznamu
Index Root, Index Allocation, Bitmap	Používány pro tvorbu záznamů adresářů (viz dále)
Security Description	Popisuje, kdo vlastní soubor a má k němu přístup.
Volume Information	Verze oddílu (pouze u metadata souboru $Volume)
Volume Name	Jméno oddílu (pouze u metadata souboru $Volume)

Atribut Attribute List

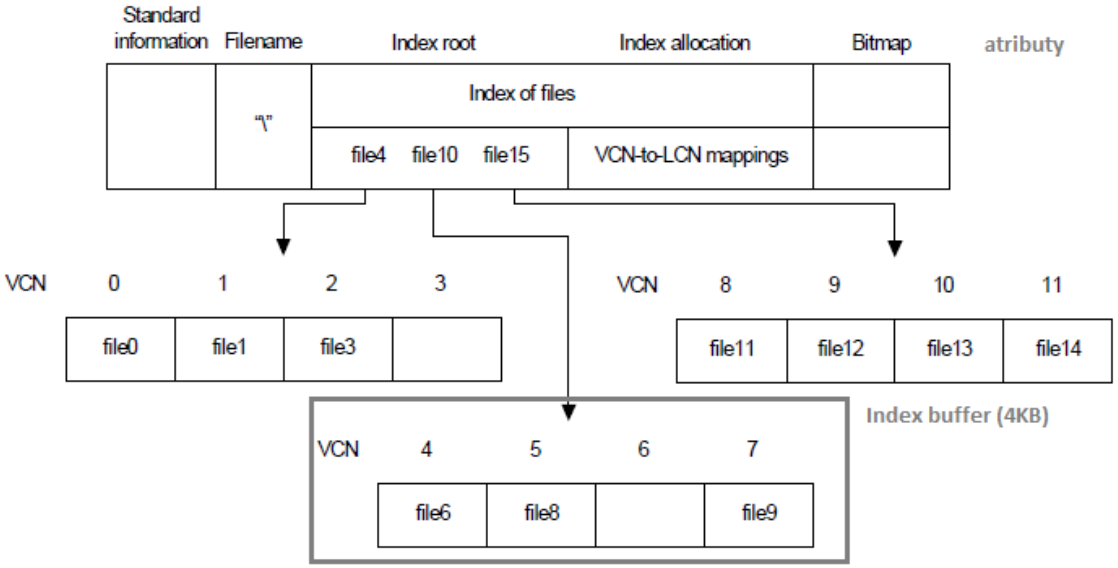


- atributy souboru se nemusí vejít do jednoho záznamu => vytvoří se další záznam
- používán také v rozsáhlých souborech, kde mapu VCN -> LCN všech extent oblastí nedokáže pojmout jeden záznam MFT
- do hlavního záznamu přidán atr. Attribute-list
  - prvek = kód atributu a ukazatel do záznamu s daty atributu

$MFT - adresář

- z pohledu NTFS adresář speciální druh souboru
- používány na rozdíl od normálního souboru atributy Index root, Index allocation, Bitmap
- Index root
  - atribut s uspořádaným seznamem souborů v malém adresáři
  - obsahuje první úroveň B stromu u rozsáhlých adresářů
  - informace o souboru v seznamu nebo stromu zahrnují:
    - název souboru
    - standardní informace – časová razítka, opravnění,...
    - File reference

$MFT – záznam adresář



- Index allocation = adresy extent oblastí s Index bufery (prvky B stromu)
- Bitmap = informace o neobsazených clusterech v Index buferech
- každý prvek Index buferu obsahuje odkaz na další Index buffer se soubory, které abecedně jsou před souborem v nadřazeném prvku

Fragmentace

- fragmentace vzniká hlavně při růstu malých souborů do velkých
  - doporučeno defragmentovat
  - fragmentace částečně redukována extent oblastmi, které jsou tvořeny skupinami po sobě jdoucích N clusterů
- s fragmentací klesá rychlost čtení souboru

  => NTFS se snaží bránit fragmentaci alespoň $MFT
  - po formátování rezervuje defaultně 12,5% oddílu kolem $MFT (tzv. MFT zóna)

Žurnálování

- změny v NTFS prováděny transakčně, žádná transakce nezůstane nedokončena
- žurnál ukládán do metadata souboru $LogFile
- cílem zajištění konzistence POUZE systémových souborů
- omezení velikosti $LogFile
  - cyklicky (typicky 5s) zapisovány tzv. checkpointy
  - cílem checkpointu označit část záznamů pro obnovu jako již nepotřebnou
- zotavení = čtení záznamů v $LogFile k poslednímu checkpointu
  1. analýza $LogFile a určení, které části oddílu potřebují zkontrolovat a opravit
  2. znovuprovedení dokončených transakcí za posledním checkpointem
  3. rollback všech nedokončených transakcí

Rozšiřující funkce

- Šifrování (NTFS 5.0+)
  - možnost šifrovat soubory nebo celé adresáře
  - z pohledu uživatele, který je autorizován, práce jako s běžnými soubory
  - u ostatních uživatelů pokus o přístup skončí chybou
  - princip
    - generování symetrického klíče a zašifrování souboru
    - šifrování klíče pomocí veřejného šif. klíče uživatele
    - uložení zašifr. klíše do speciálního pole a připojení k zašif. souboru

- Kvóty (NTFS 5.0+)
  - možnost nastavit max. prostor uživatele na konkrétním disku nebo obecně
  - reakce na překročení varováním nebo zablokováním uživatele
  - překročení monitorována a logována

Rozšiřující funkce

- Komprimované soubory (NTFS 1.2+)
  - možnost komprimovat jednotlivé soubory, adresáře, celý oddíl
  - algoritmus LZNT1, po 16 clusterech
  - s pohledu aplikací běžné soubory
    - při otevření souboru automatická dekomprese
    - při zavření souboru automatická komprese
  - omezení
    - max. velikost clusteru 4KB
    - Microsoft nedoporučuje pro soubory > 30MB
    - soubory menší než 4KB nebo již komprimované, mohou být větší po kompresi
- Sparse soubory, Volume Shadow Copy,...

Kompatibilita s OS

- Windows
  - od Windows 2000 nativní podpora
- Mac OS X
  - v10.3+ zahrnuje kompatibilitu pro čtení
- Linux
  - přes ovladač NTFS-3G možné čtení/zápis ve většině distribucí

Limity a doporučení

- názvy souborů max. 255 znaků, zakázany názvy metadata souborů
- max. velikost oddílu teoreticky $2^{64} - 1$ clusterů
  - v praxi max. 256TB při clusteru 64KB
- max. velikost souboru teoreticky 16EB
  - v praxi max. 16TB
  - ve Windows 8 max. 256TB
- max. počet souborů v oddíle přibližně 2^{32}

- pokročilé funkce si žádají výkon i prostor v oddíle navíc
  - malý a pomalý pevný disk nemusí být vhodný
  - nezapínat pokročilé funkce, pokud je nepotřebuji
- dobře si rozmyslet velikost clusteru (i vzhledem ke kompresi)
- nekomprimovat často používané soubory
- malé oddíly nižší výkon NTFS než u rozsáhlejších
- nedělit zbytečně HDD na velké možnoství oddílů

Děkuji za pozornost.