

Operačné systémy

Súborové systémy

Ing. Martin Vojtko, PhD.



2024/2025

- 1 Súborový systém
 - Súbor
 - Adresáre
- 2 Implementácia
 - Súbory
 - Adresáre
 - Zdieľané súbory
- 3 Manažment
 - Diskový priestor
 - Zálohovanie
 - Zlepšenie odozvy
- 4 Príklady FS
 - MS-DOS
 - UNIX
- 5 Zhrnutie

Uloženie dlhodobých dát

- Hlavná pamäť má obmedzenú veľkosť.
- Procesy majú svoj vyhradený adresný priestor.
 - Tu sa môžu dočasne ukladať dáta ako výsledky spracovaní.
 - Po skončení procesu sú neuložené dáta stratené.
 - Adresný priestor procesu nie je vhodné úložisko dát.
- Viac inštancií programu potrebuje pracovať s rovnakými dátami.
- ❶ Musíme mať možnosť uložiť obrovské množstvo dát.
- ❷ Dáta musia prežiť smrť procesu.
- ❸ Dáta sú prístupné viacerým procesom súčasne.
- Súčasné disky nám toto umožňujú pomocou operácií Read/Write Block.

Súborový systém

Súborový systém

File (Súbor)

je abstrakcia OS, ktorá umožňuje prácu s dátami používateľa bez nutnosti bližšie poznať zariadenie na, ktorom sú dáta uložené. Súbor je vlastne štruktúra, ktorá obsahuje informácie o blokoch disku, ktoré mu patria.

Directory (Adresár)

je vo všeobecnosti špeciálny súbor, ktorý obsahuje informáciu o všetkých súboroch a iných adresároch, ktoré sú pod nim vedené.

File System (Súborový systém)

je štruktúra adresárov s uloženými súbormi. Z pohľadu OS je File System abstrakciou so všeobecne známym API skrývajúcim zložitý HW.

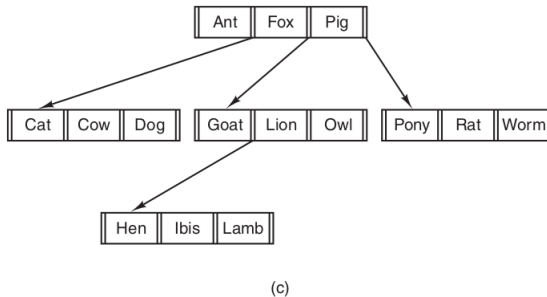
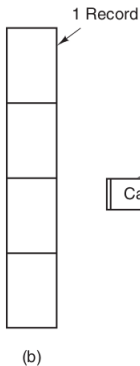
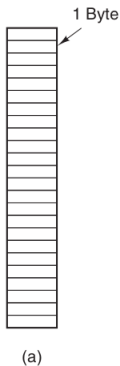
Súbor - meno

- Plní dôležitú úlohu pri identifikovaní.
- Väčšina OS umožňuje použiť:
 - písmená abecedy,
 - čísla
 - špeciálne znaky.
- Každý OS definuje znaky, ktoré nemôžu byť súčasťou mena súboru.
- Niektoré OS nerozlišujú medzi malými a veľkými písmenami abecedy.
- Bežné použitie ASCII, utf-8 alebo unicode.
- Niektoré OS identifikujú aj typ súboru príponou.
 - v UNIX sa prípona považuje za súčasť názvu súboru.
 - Windows berie ohľad pri práci so súborom aj na príponu.

Súbor - štruktúra

- Sekvencia Byte-ov. (a)
 - OS nediktuje čo je obsahom súboru.
 - Vybraná aplikácia parsuje obsah.
 - Prístup WINDOWS, UNIX...
- Sekvencia záznamov. (b)
 - Záznam fixnej dĺžky definuje minimálnu jednotku súboru.
 - V časoch 80 znakových riadkov bol súbor bežne definovaný ako séria 80 znakových riadkov.
- Strom. (c)
 - Namiesto záznamov uložených jeden za druhým sú uložené v strome.
 - Každý záznam má svoj kľúč.

Súbor - štruktúra



Súbor - druhy súborov

- Regulárne súbory
 - ASCII súbory - čitateľné tak ako sú. Umožňujú jednoduché spracovanie.
 - Binárne súbory - obsah zobraziteľný len s aplikáciou, určenou na čítanie daného súboru.
- Adresáre
- Špeciálne súbory
 - Znakové súbory - modelovanie znakových zariadení (tlačiareň, terminál)
 - Blokové súbory - modelovanie blokových zariadení (disk)

Súbor - atribúty

Attribute	Meaning
Protection	Who can access the file and in what way
Password	Password needed to access the file
Creator	ID of the person who created the file
Owner	Current owner
Read-only flag	0 for read/write; 1 for read only
Hidden flag	0 for normal; 1 for do not display in listings
System flag	0 for normal files; 1 for system file
Archive flag	0 for has been backed up; 1 for needs to be backed up
ASCII/binary flag	0 for ASCII file; 1 for binary file
Random access flag	0 for sequential access only; 1 for random access

1

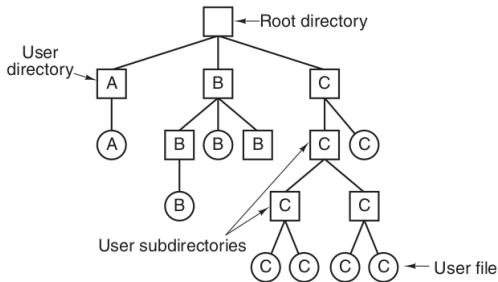
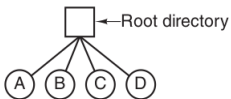
¹ďalšie príklady nájdete v MOS od Tanenbauma

Súbor - operácie/systémové volania

- Create() a Delete()
- Open() a Close()
- Read(), Write(), Append() a Seek()
- GetAttributes() a SetAttributes()
- Rename()

Adresáre

- Jedno-úrovňové systémy
 - FS má jeden root v ktorom sú uložené všetky súbory.
 - Jednoduchá správa adresárov :)
- Hierarchické systémy
 - FS obsahuje množstvo adresárov tvoriacich štruktúru.
 - Komplexná správa FS. Prívetivejšia štruktúra dát používateľov.



Adresáre - cesta k súboru (Path name)

absolútna cesta

je cesta k súboru od koreňa FS cez všetky adresáre, ktoré možno označiť za predchodcov súboru.

pracovný adresár (working directory)

je aktuálny adresár, v ktorom sa používateľ alebo program nachádza. Tento adresár je súčasťou premenných prostredia.

relatívna cesta

je cesta k súboru od pracovného adresára.

. a ..

vo väčšine OS s hierarchickým FS existujú v každom adresári špeciálne adresáre definujúce alias pre aktuálny adresár "." a alias pre rodičovský adresár "..". V koreňovom adresári "." referencuje samého seba.

Adresáre - operácie/systémové volania

- CreateDir() a DeleteDir()
- OpenDir() a CloseDir()
- ReadDir()
- Link() a Unlink()
- Rename()

Linky

- Najjednoduchšia forma FS je strom.
- Ak chceme zamedziť duplikácií súborov je nutné aby bolo možné vo viacerých adresároch referencovať tie isté súbory.
- Z FS vznikne orientovaný acyklický graf.

hard link

odkaz v adresári FS na súbor, ktorému zdvíham počet referencií. Pri mazaní takéhoto súboru musím zmazať všetky inštancie.

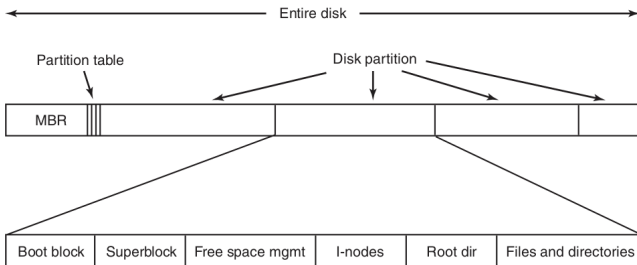
soft/symbolic link

je súbor, ktorého obsahom je cesta k inému existujúcemu súboru. Ak zmažem symbolickú linku neovplyvníme referencovaný súbor. Ak zmažem súbor tak symbolická linka odkazuje na neexistujúci súbor.

Implementácia

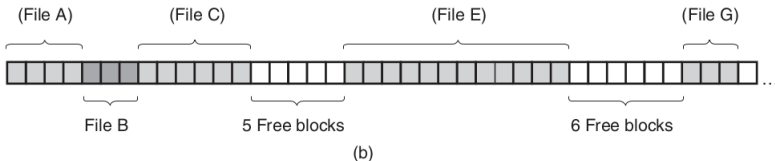
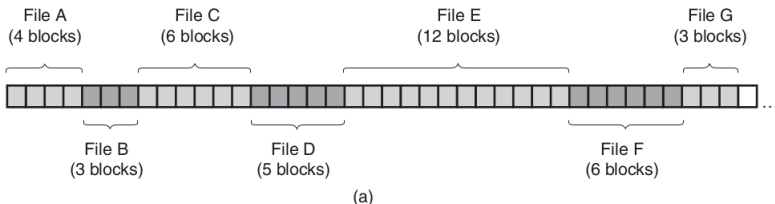
FS layout

- Bežné rozloženie disku:
 - MBR - Master Boot Record - sector 0 disku používaný na boot OS. Obsahuje zavádzací program.
 - Partície - logické diskové jednotky, ktoré môžu obsahovať FS.
- Bežné rozloženie FS:
 - Bootblock - je to čo sa snaží v partíciách prečítať zavádzač.
 - Superblock - informácie o FS (magic number, počet blokov FS...)



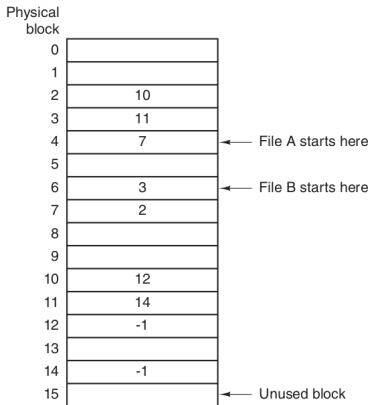
Uloženie súboru - Kontinuálny sled

- externá fragmentácia
- ideálne len na write-once úložiská CD/DVD



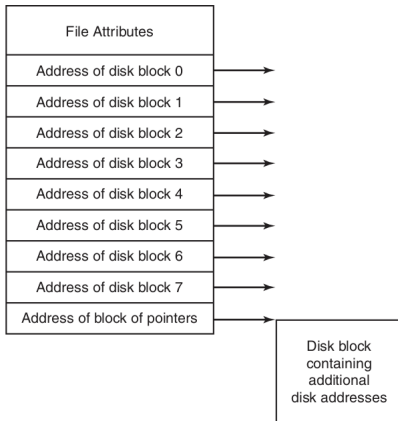
Uloženie súboru - File Allocation Table (FAT)

- Bežne používaný FS (FAT-32).
- Celá FAT sa pri inicializácii FS načíta do hlavnej pamäte.



Uloženie súboru - i-node

- index-node je štruktúra udržiavajúca informácie o blokoch súboru.
- v pamäti mám iba i-nody používaných súborov.

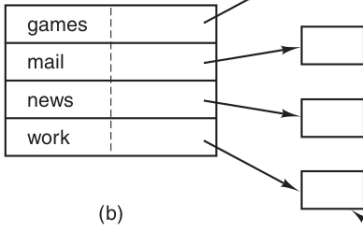


Adresáre

- Adresár obsahuje záznam pre každý uložený súbor.
- Súbor reprezentujeme menom, atribútmi a zoznamom blokov.
- Meno súboru je uložené v štruktúre adresára.
- Atribúty a ukazovateľ na zoznam blokov môžeme:
 - uložiť priamo v štruktúre adresára. (a)
 - uložiť nepriamo referenciou na i-node. (b)

games	attributes
mail	attributes
news	attributes
work	attributes

(a)



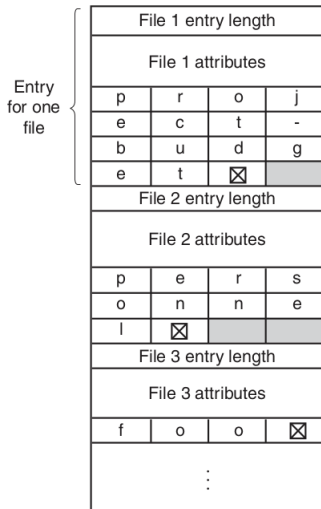
(b)

Data structure
containing the
attributes

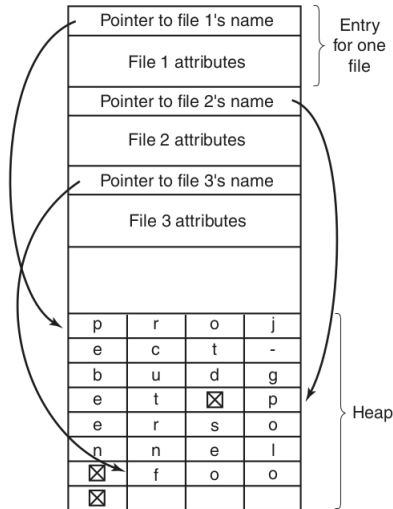
Adresáre - uloženie mien súborov

- V minulosti bol názov súboru limitovaný na 8+3 resp. 14 znakov.
- Meno súboru bolo uložené v poli fixnej dĺžky.
- V súčasnosti môžu byť názvy súborov dlhšie (255).
- Meno súboru je uložené v poli variabilnej dĺžky:
 - ako súčasť záznamu súboru. (a)
 - na halde (heap) s odkazom na meno v zázname. (b)

Adresáre - uloženie mien súborov



(a)



(b)

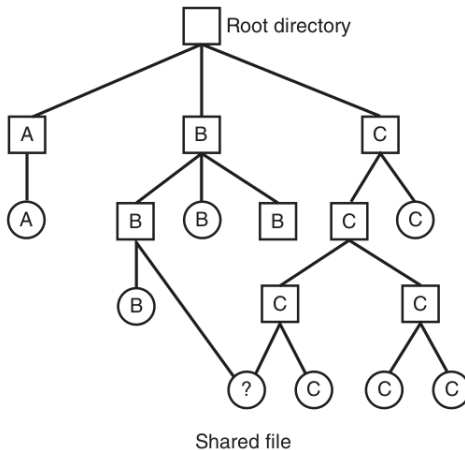
Adresáre - uloženie mien súborov

- ako súčasť záznamu súboru (a)
 - záznamy rôznej veľkosti spôsobujú fragmentáciu.
 - prehľadávanie obsahu adresára je komplexný.
 - pravdepodobné množstvo page faults.
- na halde (heap) s odkazom na meno v zázname. (b)
 - záznamy majú rovnakú veľkosť, práca je jednoduchšia.
 - heap uložený na koniec adresného priestoru adresára.
 - prehľadávanie obsahu adresára vyžaduje prechádzanie zoznamom.
- uloženie položiek v hash tabuľke môže urýchliť prehľadávanie.

Zdieľané súbory

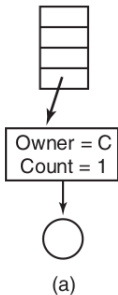
- FS umožňujú vytváranie tzv. liniek, ktoré redukujú duplikácie.
- Linky menia FS zo stromu na orientovaný acyklický graf.
- Hard linka
 - všetky inštancie záznamu o súbore sú rovnocenné a ukazujú na rovnaký i-node.
 - i-node udržiava ref. count (počet odkazov).
 - zmazanie v jednom adresári nesmie ovplyvniť stav v druhom.
 - ak zmažeme všetky inštancie záznamov môžeme zmazať aj súbor.
- Soft linka
 - existuje len jedna inštancia súboru v jednom adresári.
 - existuje viac špeciálnych súborov, ktoré obsahujú len cestu k originálnemu súboru.

Zdieľané súbory

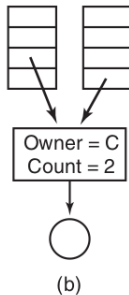


Zdieľané súbory - hard link

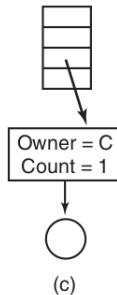
C's directory



B's directory C's directory



B's directory



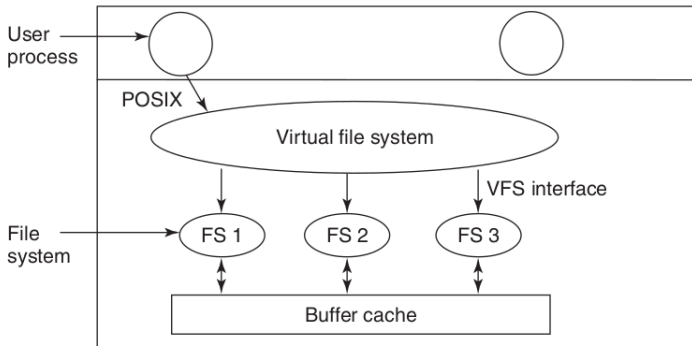
Journaling

- Pri práci s výpočtovým systémom môže kedykoľvek dôjsť k poruche.
- FS sa pri poruche môže dostať do nekonzistentného stavu.
 - Napr. zmazanie súboru obnáša:
 - zmazanie súboru z adresára.
 - uvoľnenie i-node.
 - uvoľnenie všetkých blokov disku.
- Bežný FS preto udržiava žurnál vykonaných operácií.
 - pred vykonaním operácie sa uloží záznam o všetkých krokoch operácie na disk.
 - vykoná sa operácia.
 - ak bola operácia úspešná zmaže sa záznam.
- Pri bootovaní systému sa kontroluje, či je žurnál prázdny. Ak žurnál nie je prázdny, musí sa vykonať oprava.

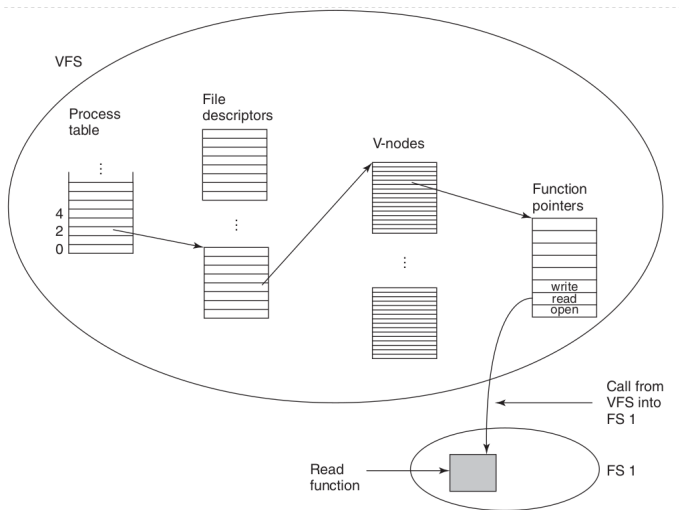
Virtual FS

- Existuje množstvo druhov FS.
- Ako sa dá zabezpečiť ich ko-existencia?
- Windows to vyriešil pomocou drive letter (C:, D:)
 - Výhoda je jednoduchosť implementácie.
 - každý FS má vlastný strom.
- Linux sa vydal cestou virtuálneho FS.
 - Existuje root FS do ktorého vieme pomocou "mount()" linkovať iné druhy FS.
 - Virtuálny FS definuje jednotné rozhranie, ktoré ostatné FS musia podporovať.
 - Virtual FS je definovaný ako POSIX štandard.
 - Používateľ ani len netuší, že pri prechádzaní stromu mení FS.

Virtual FS



Virtual FS



Manažment

Veľkosť bloku

- Veľké bloky:
 - súbory majú málo blokov - redukuje sa čas prečítania súboru.
 - malé súbory spôsobujú plytvanie priestoru disku.
- Malé bloky:
 - súbory zaberajú množstvo blokov - čas čítania súboru rastie (seek()).
 - malé súbory neplytvajú pamäť.
- Stanovenie veľkosti bloku závisí aj od určenia výpočtového systému.
 - dátové úložisko filmov (GB súbory)
 - archív dokumentov resp. faktúr (KB - MB súbory)

Veľkosť bloku - štúdia

Percento súborov menších ako veľkosť bloku vybranej veľkosti.

Length	VU 1984	VU 2005	Web
1	1.79	1.38	6.67
2	1.88	1.53	7.67
4	2.01	1.65	8.33
8	2.31	1.80	11.30
16	3.32	2.15	11.46
32	5.13	3.15	12.33
64	8.71	4.98	26.10
128	14.73	8.03	28.49
256	23.09	13.29	32.10
512	34.44	20.62	39.94
1 KB	48.05	30.91	47.82
2 KB	60.87	46.09	59.44
4 KB	75.31	59.13	70.64
8 KB	84.97	69.96	79.69

Length	VU 1984	VU 2005	Web
16 KB	92.53	78.92	86.79
32 KB	97.21	85.87	91.65
64 KB	99.18	90.84	94.80
128 KB	99.84	93.73	96.93
256 KB	99.96	96.12	98.48
512 KB	100.00	97.73	98.99
1 MB	100.00	98.87	99.62
2 MB	100.00	99.44	99.80
4 MB	100.00	99.71	99.87
8 MB	100.00	99.86	99.94
16 MB	100.00	99.94	99.97
32 MB	100.00	99.97	99.99
64 MB	100.00	99.99	99.99
128 MB	100.00	99.99	100.00

Veľkosť bloku - štúdia

- Pri veľkosti bloku 4KB sa 70% súborov zmestí do jedného bloku.
- Na druhú stranu 93% blokov je obsadených 10% najväčších súborov (videá)
- Takže napriek tomu že 70% súborov plytvá pamäťou v globálnom pohľade maximálne 7% blokov nie je plných.
- veľkosť bloku určuje aj rýchlosť čítania súboru.

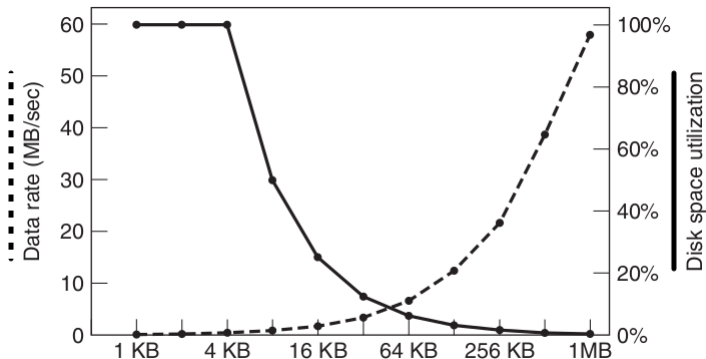
Čas čítania bloku v HDD

majme t_s priemerný čas vyhľadania stopy, t_r čas jedného otočenia disku, k veľkosť bloku v B a s veľkosť stopy v B. Potom čas prečítania bloku bude:

$$t = t_s + \frac{t_r}{2} + \frac{k}{s} t_r$$

Veľkosť bloku - štúdia

Rýchlosť čítania a využitia disku (1MB) pri súboroch veľkosti 4KB a rôznych veľkostiach blokov.



Zálohovanie

- Pravidelné odkladanie práce na disk zabezpečuje, že výstupy našich programov sa nestratia ak vypadne napájanie výpočtového systému.
- Ale ak sa niečo stane disku tak stratíme všetky dáta.
- Koľký z vás si robia zálohy?
 - Ak sa pokazí počítač tak jeho cena je niekoľko rádov nižšia ako cena dát uložených na ňom.
 - Na PC máte uložené svoje spomienky, dokumenty, heslá, virtuálnu menu...
- zálohovanie umožňuje obnovu dát:
 - po katastrofe (blesk, vytopenie, krádež, nepříčetný spolubývajúci...)
 - po katastrofe medzi stoličkou a klávesnicou (alkohol v krvi, ranný šialok čaju na klávesnici, hodina spánku mi stačí, materiály na OS už nebudem potrebovať...)

Zálohovanie - čo zálohovať?

- Zálohovanie disku môže trvať veľmi dlho. Preto je dobré zálohovať iba to potrebné
 - Nainštalované programy vieme znova stiahnuť z CD/DVD/internetu.
 - Vlastné programy môžeme zálohovať na git repozitári
 - Nikdy nezálohujeme /dev adresár!!!!
- Je dobré udržiavať dôležité dáta na jednom mieste.

Zálohovanie - stratégie zálohovania a obnova

- recycle bin
 - Windows zmazané súbory presúva do koša. Nie je to záloha ale ochrana.
- kompletne zálohovanie
 - Všetky adresáre, ktoré zálohujem skopírujem na iný disk.
 - Takéto zálohy sa robia na mesačnej báze.
- prírastkové zálohovanie
 - Zálohovací systém ukladá iba zmeny voči poslednej kompletnej zálohe.
 - Inkrementálne zálohy sa robia na dennej báze.
- Obnova býva zväčša veľmi zdĺhavá.
- Pri úplnej obnove sa najprv načíta posledná kompletná záloha FS.
- Následne sa postupne aplikujú všetky prírastkové zálohy.

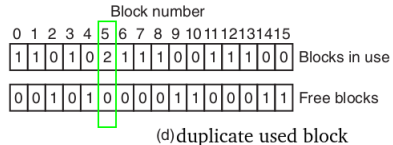
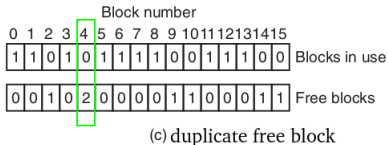
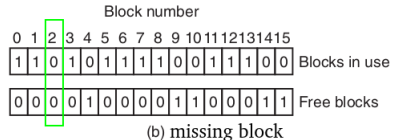
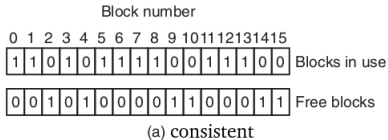
Zálohovanie - implementácia

- Fyzické zálohy
 - Ukladá sa fyzický obsah disku blok za blokom.
 - Nezáleží na FS.
 - Takýto zálohovací softvér je veľmi jednoduchý.
 - Kopíruje sa všetko aj voľné bloky a poškodené bloky.
- Logické zálohy
 - Zálohovací algoritmus rozumie FS.
 - Pomalšie zálohovanie.
 - Jednoduché nastavenie granularity zálohovania.
 - Umožňuje inkrementálne zálohovanie.

Konzistentnosť FS

- Počas života FS vznikne množstvo situácií, kedy dôjde k chybe.
- väčšina FS je vybavená kontrolným mechanizmom (fsck, sfc)
- Bloková konzistentnosť - počíta sa počet referencií každého bloku.
- Súborová konzistentnosť - počíta sa počet referencií súboru a porovnáva sa s počtom referencií v i-node.

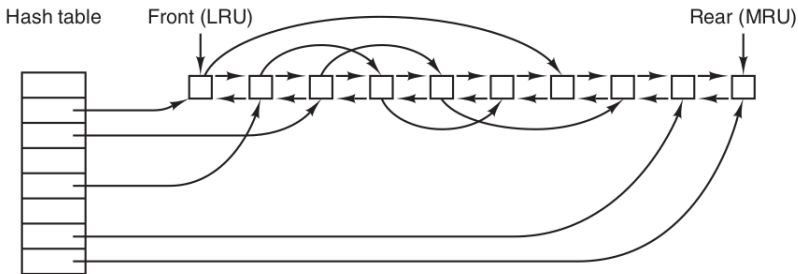
Konzistentnosť FS - bloková



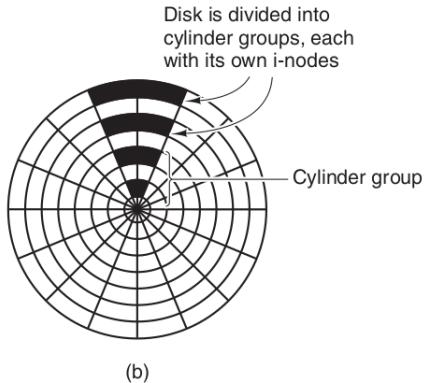
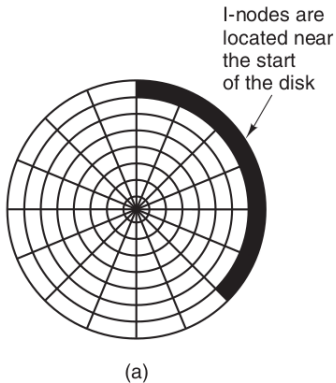
Zlepšenie odozvy FS

- Block cache:
 - Zmenšenie čítaní disku výrazne urýchli systém.
 - Často používané bloky disku môžeme udržiavať v hlavnej pamäti.
- Block read ahead:
 - Pri čítaní blokov súboru čítame okrem bloku k aj blok $k+1$.
 - aplikovateľné iba pre sekvenčné súbory.
- Disk-Arm Motion Reduction.
 - i-node FS vyžadujú čítanie minimálne dvoch blokov.
 - Ukladanie blokov súboru a i-node v rámci jedného cylindra.
 - Alebo ukladanie i-node blokov do stredných cylindrov disku.
- File Path caching:
 - Parsovanie cesty k súboru je zdĺhavý proces.
 - Už raz použité cesty a k nim prislúchajúce i-nodes môžeme čítať z cache.

Zlepšenie odozvy FS - caching



Zlepšenie odozvy FS - arm motion



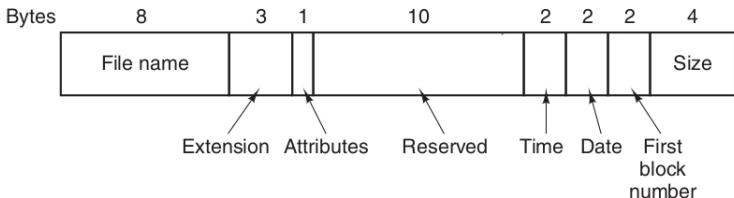
Príklady FS

MS-DOS

- Jednoduchý, úspešný, široko rozšírený.
- Používaný na menšie úložiská ako sú fotoaparáty, USB disk, vnorené systémy.
- Vychádza z neho aj FAT-32.
- Je používannejší ako NTFS.

MS-DOS - položka adresára

- meno súboru maximálne 8+3 znaky.
- čas zmeny 5b - s, 6b min, 5b h
- datum zmeny 5b - d, 4b - m, 7b - r (od 1980 do 2107)
- veľkosť súboru max 2GB
- maximálne 65K blokov - toľko má záznamov FAT tabuľka.



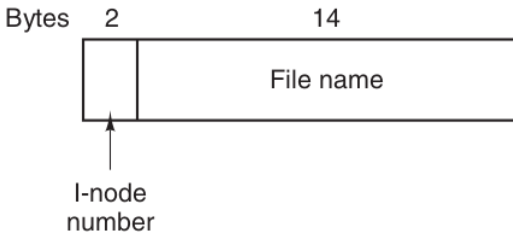
MS-DOS

- rôzne verzie FS existujú podľa veľkosti adresnej zbernice.
- FAT-12, FAT-16, FAT-32(28), exFAT
- FAT-32 už podporuje mená dlhé 255 znakov.

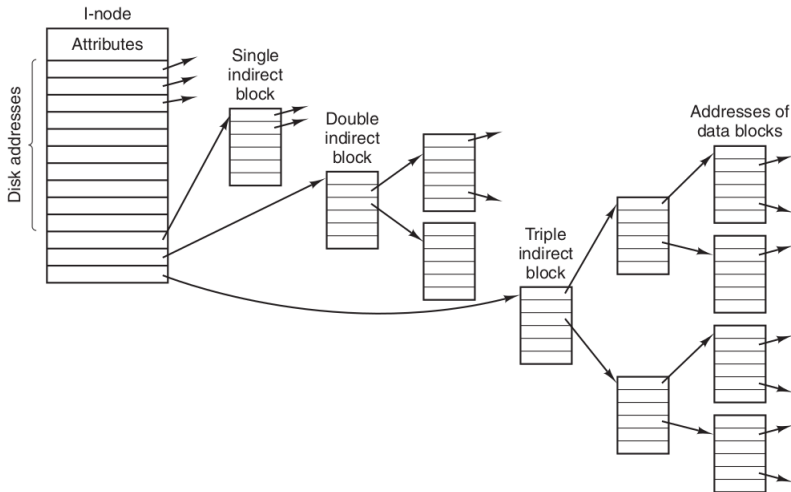
Block size	FAT-12	FAT-16	FAT-32
0.5 KB	2 MB		
1 KB	4 MB		
2 KB	8 MB	128 MB	
4 KB	16 MB	256 MB	1 TB
8 KB		512 MB	2 TB
16 KB		1024 MB	2 TB
32 KB		2048 MB	2 TB

UNIX - V7

- meno súboru max 14 znakov. Zakázané znaky sú / a NUL
- i-node number - max 65K súborov.
 - file-size, owner, group, protection, dir. ref. count.
 - creation, last access, modify time
 - voľné i-nodes sú udržiavané v zozname.



UNIX - V7 i-node



UNIX - V7 /usr/ast/mbox

Root directory

1	.
1	..
4	bin
7	dev
14	lib
9	etc
6	usr
8	tmp

Looking up
usr yields
i-node 6

I-node 6
is for /usr

Mode size times
132

I-node 6
says that
/usr is in
block 132

Block 132
is /usr
directory

6	.
1	..
19	dick
30	erik
51	jim
26	ast
45	bal

/usr/ast
is i-node
26

I-node 26
is for
/usr/ast

Mode size times
406

I-node 26
says that
/usr/ast is in
block 406

Block 406
is /usr/ast
directory

26	.
6	..
64	grants
92	books
60	mbox
81	minix
17	src

/usr/ast/mbox
is i-node
60

Zhrnutie

Zhrnutie

- Bežne používané FS môžeme rozdeliť do troch kategórií
 - súborové systémy trvalých nosičov dát.
 - FAT súborové systémy používané najmä vo vnorených systémoch.
 - i-node súborové systémy používané v serverových a PC aplikáciách.
- Krása súborových systémov je, že vás odbremeňujú od komplexnosti diskov.
- Je dôležité si chrániť vaše dáta
 - zálohujte na externý disk.
 - využite cloud služby - overte si službu pred tým než jej zveríte vaše dáta.
 - používajte versioning nástroje ako je git.

.

Čo robiť do ďalšej prednášky

- Prečítať kapitolu 5. z Tanenbauma.