41. METADATA

Metadata – raison informace o souborech
- jsou to data o datech
- nesou informace o pojmenování souborů, jejich organizaci a další

Umístění – přímo v adresářích
- ve speciálních strukturách na vyhrazeném místě na paměti
- ty speciální struktury mohou být I-uzly, B+ stromy nebo třeba MFT (Master File Table) v NTFS

Použití metadat
- zpřístupnění – hledají se informace o datech/souboru na základě jeho jména
- čtení a zápis – k tomu se používají deskriptory souborů (v podstatě také metadata protože popisují soubory)
- manipulace se soubory – v podstatě se jedná o úpravu a změnu metadat souborů

Soubor – abstraktní jednotka která obsahuje data a je jednoznačně identifikována jménem a třeba I-uzlem nebo B+ stromem atd..

Jméno souboru
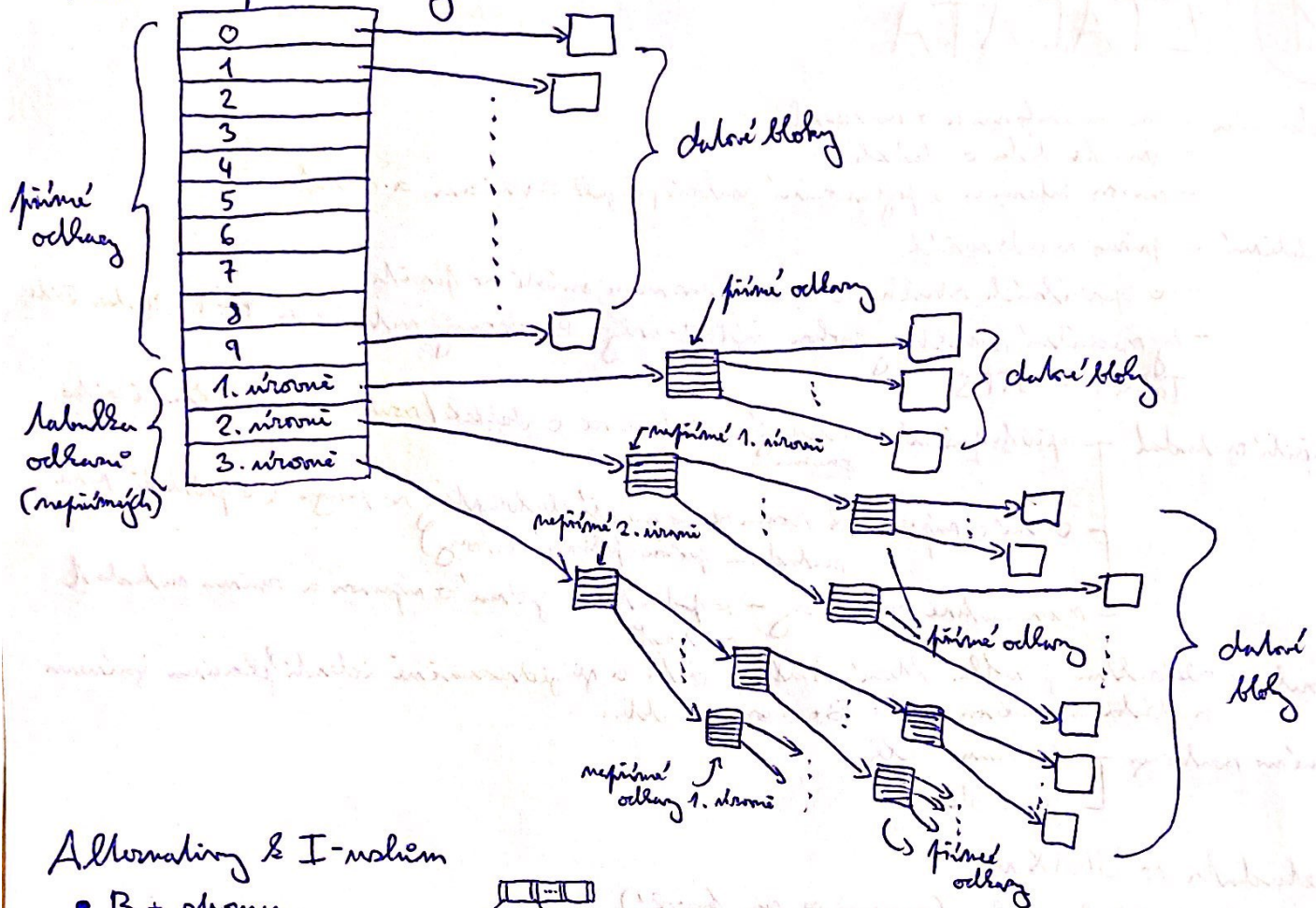- proměnné délky
- pevné délky

Metadata v UNIXu

Obsah logického disku (mapují se na fyzický)

Boot block
Super block
I-nodes 1..32
I-nodes 33-64

....
I-nodes n
Data block
Data block
....
Data block

- Boot block – data OS potřebná k nabootování
- Super block – data o souborovém systému – typ (ext, ext2,...), počet I-uzlů, kapacita, volná kapacita, volné I-uzly,...
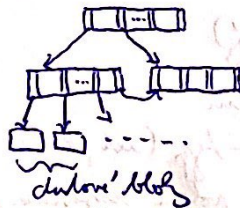- Tabulka I-uzlů – tabulka popisů souborů
- datové bloky

I-uzel
- datová struktura popisující soubor v UNIXových systémech
- obsahuje metadata a občas i data – pro malé soubory nebo pro symlinky
- obsahuje: stav uzlu (plný/prázdný)
  - vlastníka souboru a skupinu
  - práva k souboru (0644, r--rw, chmod,..)
  - typ souboru – adresář, obyčejný, sokel,...
  - velikost souboru v bajtech (B)
  - čas poslední modifikace souboru, čas poslední modifikace I-uzlu, čas posledního přístupu do souboru
  - několik přímých odkazů na datové bloky (10-12)
  - tabulka nepřímých odkazů na datové bloky – 1. úrovně, 2. úrovně a 3. úr.

①

Přímé a nepřímé odkazy I-uzlu

přímé odkazy

tabulka odkazů (nepřímých)

0
1
2
3
4
5
6
7
8
9
1. úrovně
2. úrovně
3. úrovně

datové bloky

přímé odkazy

datové bloky

nepřímé 1. úrovně

nepřímé 2. úrovně

přímé odkazy

nepřímé odkazy 1. úrovně

přímé odkazy

datové bloky

Alternativy k I-uzlům

• B+ stromy

datový blok

• Adresáře – nemusí to být úplně alternativy ale týká se i se spojením s I-uzly
  – adresáře jsou buď speciální struktury a musí se o nimi také tak [resp. pracovat]
  a nebo to jsou typy souborů a tedy jsou reprezentovány I-uzly a jménem atd...
  – adresář jako soubor obsahuje dvojice
  (jméno souboru, identifikátor souboru)
  ↳ jž třeba právě jen I-uzel

< jednoúrovňový adresář
  víceúrovňový adresář – třeba strom

Organizace volného prostoru

  – v podstatě to jsou také metadata která určují zda ji daný prostor volný nebo
  ne
  – bitová mapa
  – 1 bit pro každý datový blok
  – bit 0 = volný x bit 1 plný datový blok
  – pomocí maskování se hledají souvislá volná místa / prostory

Odolnost souborového systému vůči výpadku

- příkladem situace kdy může nastat výpadek je smazání souboru a vzniknout při něm nekonzistence
- smazání souboru má několik kroků které musí jít synchronně po sobě a pokud jdou asynchroně a nastane výpadek tak vznikají nekonzistence v metadatech a neuvnitřelný stav metadat (mohou jít i asynchroně a ne jen synchroně ale při výpadku se nebe oslaví)
  - smazání souboru je totiž asynchronní úprava jejich metadat
=> řešení:
  1) synchronní zapisovat metadata v plně diném přístupu => pomalé
  2) soft metadata update — sledují se závislosti metadat (musí - graf závislostí)
     — zaručují konzistenci metadat na disku
  3) žurnalizační operování — vše se loguje a při výpadku se provedou operace z logu které nebyly zapsány na disk před výpadkem