

---

# Administrace diskového subsystému, systém souborů, swap

Úvod do administrace operačních systémů

Jan Žďárek

České vysoké učení technické v Praze  
Katedra počítačů

(Program STM – kód Y36UAD)





# Úvod

---

Motivace:

Bez diskového úložného prostoru se nejde obejít.



# Úvod

---

Motivace:

Bez diskového úložného prostoru se nejde obejít. . . ?



# Úvod

---

Motivace:

Bez diskového úložného prostoru se nejde obejít. . . ?  
(Embedded devices, PROM, Flash, net. . . )



# Úvod

---

Motivace:

Bez diskového úložného prostoru se nejde obejít. . . ?  
(Embedded devices, PROM, Flash, net. . . )

1. IDE (ATA), SATA,
2. SCSI,
3. ??? (FireWire IEEE1394, FibreChannel, USB, . . . ).
4. Souborový systém.
5. Zařízení (soubor).



# Přidání disku (1)

---

## HW:

- Fyzická instalace a připojení na sběrnici v počítači.
- (Formátování disku).

## SW:

- Vytvoření speciálního souboru (*blokového zařízení*), reprezentujícího nový disk v OS.
- Partitioning, vytvoření logických disků (partitions, slices).
- Vytvoření souborových systémů na logických discích.
- Konfigurace připojování.
- Nastavení swap partitions.



# Přidání disku (2)

HW: Formátování disku.

- Výrobci často udávají velikost disků v neformátované kapacitě.
- Fyzický formát zapisuje na disk časovací značky a další pomocné informace – vytvoří stopy.
- Část disku vyhrazena pro on-the-fly náhradu vadných sektorů (HW, řadič).
- ATA disky předformátované, provedeno vyššími proudy, na celou dobu životnosti disku.
- ATA disky (obvykle) nejde přeformátovat mimo výrobní linku.
- SCSI disky s formátováním počítají, součást SCSI příkazů.
- SCSI disky se dodávají také předformátované a formátování je zbytečné.



# Logická reprezentace zařízení

Idea: Mnoho různých zařízení může být připojeno do jednoho adresářového stromu.

Speciální soubory – *zařízení*.

► Typ, major a minor číslo.

## ● Znaková

- `/dev/ttyS0`
- `/devices/ebus@800/serial@0,c2c000:a`

## ● Bloková

- `/dev/sda1`
- `/devices/pci@7c0/pci@0/pci@8/scsi@2/sd@0,0:a`
- `/devices/pci@7c0/pci@0/pci@8/scsi@2/sd@0,0:a,raw`
- `/dev/dsk/c0t0d0s0`
- `/dev/rdisk/c0t0d0s0`





# Přidání disku (1)

---

SW: Vytvoření speciálního souboru (*blokového zařízení*), reprezentujícího nový disk v OS.

mknod

MAKEDEV

udev

ENOENT: No such file or directory – chybí zařízení,

ENODEV: No such device – kernelu chybí ovladač.



# Přidání disku (2)

---

SW: Partitioning, vytvoření logických disků (partitions, slices).

`fdisk` *blokové zařízení*

(Solaris/SPARC: `format`)



# Systémy souborů

---

- Diskové
  - s5, minix, ext;
  - UFS, BSD Fast Filesystem, ext2;
  - vxfs, ext3...
  - ZFS, ext4 (?)...
- Síťové
  - NFS, SambaFS...
- PseudoFS
  - procfs, fdfs...
  - tmpfs...



# Přidání disku (3)

---

SW: Vytvoření souborových systémů na logických discích.

`mkfs`

`mkfs.ext2`

`mkfs.vfat` – umožňuje nastavování parametrů podle skutečných možností FAT, zejména FAT-32. Nativní prostředky výrobce Windows jsou omezené.

`mkfs...`



# Přidání disku (4)

---

SW: Konfigurace připojování při startupu systému.

/etc/fstab

```

#(fs_spec) (fs_file) (fs_vfstype) (fs_mntops) (fs_freq) (fs_passno)
# <fs>      <mountpoint>      <type>          <opts>          <dump/pass>
/dev/sda1    /                ext3            rw              1 1
/dev/sdb1    none            swap            swap            0 0
/dev/hda1    /home           ext3    auto,rw,nodev,nosuid  1 2
/dev/hda2    none            swap            defaults        0 0
/dev/cdrom   /mnt/cdrom      udf,iso9660     noauto,owner,ro 0 0
/dev/fd0     /mnt/floppy     auto            noauto,owner    0 0
```



# Přidání disku – připojení

---

mount

mount *[options] blokové\_zařízení kam*

**Odpojení:** umount

umount *blokové\_zařízení*

umount *cesta*



# Přidání disku – swap

---

**SW:** Nastavení swap partitions.

`mkswap`

`mkswap -c` (check badblocks)

`swapon`

**Editace** `/etc/fstab`.

`swapon -a` – **vše** z `fstab`, kromě již používaných a označených *noauto*.

`swapon -p prio - fstab: pri=prio, 0–32767.`

**Vypnutí:**

`swapoff`

`swapoff -a`



# System souborů – superblok

---

## Statické informace:

- typ FS (*magic*, ext2: 0xEF53),
- velikost FS,
- počty i-nodes a datových bloků,
- kořenový i-node,
- počet rezervovaných bloků...

## Dynamické informace:

- počty volných i-nodes a datových bloků,
- stav FS (mounted/unmounted (clean))...

`tune2fs -l`

(Solaris: `fstyp -v`)





# System souborů – skupina bloků

---

UFS: cylinder group

Deskriptor (superblok) skupiny bloků – adresy bloků:

- bitmapa alokovaných datových bloků,
- bitmapa alokovaných i-nodes,
- počet volných i-nodes a datových bloků,
- počet adresářů ve skupině.



# Systém souborů – i-node

---

- Vlastník (UID/GID),
- velikost v B,
- velikost v alokovaných blocích,
- počet hardlinků,
- práva,
- typ,
- čas
  - posledního přístupu (*a-time*),
  - poslední modifikace dat (*m-time*),
  - vytvoření/poslední modifikace i-node (*c-time*).
- 10–12 čísel datových bloků,
- 1 datový blok nepřímé adresace,
- 1 datový blok dvojité nepřímé adresace,
- 1 datový blok trojitě nepřímé adresace.



# Kontrola a opravy FS

---

Přes všechna opatření se může stát, že dojde k poškození FS. Výpadek napájení, nahodilý reset – neohrozí konsistenci žurnálovacích FS.

Žurnálovací FS nemůže zachránit data! Nezapsaná data v cache jsou nenávratně ztracena, na disku bude v nejlepším případě stará verze dat.

Existují i chyby, proti nimž není obrany – chyby v jádře: nejen v ovladači FS, ale v libovolém kódu jádra (neotestované moduly!). Chybná manipulace s datovými strukturami jádra.

`fsck`

`fsck.ext2; fsck.vfat; ...`



# fsck (1)

---

Pokud `fsck` ze superbloku zjistí, že FS byl čistě odpojen (bit *clean*), kontrolu neprovede.

`fsck -f`

Co `fsck` hledá a umí opravit:

- obsazené, ale neodkazované i-nody,
- realitě neodpovídající čítače odkazů (LC),
- nepoužité datové bloky chybějící v mapách volných bloků,
- datové bloky označené jako volné, ale použité pro data,
- nesprávné hodnoty pomocných informací v superbloku.



## fsck (2)

---

Co `fsck` hledá a neumí sám opravit:

- bloky odkazované ze dvou souborů,
- odkazy na bloky mimo FS,
- nižší než skutečný počet odkazů v LC,
- neevidované bloky,
- adresář odkazující na nealokovaný i-node,
- chyby v položkách i-nodů (formát).

Každý logický disk (UFS, ext2, ...): adresář `lost+found`.