

# Linuxový kernel v posledních letech

Jiří Kosina

<jiri.kosina@suse.com>

a

Jan "Yenya" Kasprzak

<kas@fi.muni.cz>

<http://www.fi.muni.cz/~kas/>

SUSE Labs, Masarykova univerzita

XXXI. konference EurOpen.CZ

# Obsah přednášky

- 1** Vývojový model Linuxu
- 2** Co je nového v kernelu?

# Anketa na úvod



## Otázka:

Kdo v posledním týdnu **nepoužil** Linux?

# Anketa na úvod



## Otázka:

Kdo v posledním týdnu **nepoužil** Google?

# Anketa na úvod



## Otázka:

Kdo v posledním týdnu **ne**použil Linux?

# Vývojový model Linuxu

## 1 Vývojový model Linuxu

## 2 Co je nového v kernelu?

# Linux je velký projekt

- > 1000 aktivních vývojářů
- Jak zajistit kvalitu kódu?
- Jak zajistit propustnost?



# Linux je velký projekt

- > 1000 aktivních vývojářů
- Jak zajistit **kvalitu kódu**?
- Jak zajistit **propustnost**?





# Linux je velký projekt

- > 1000 aktivních vývojářů
- Jak zajistit kvalitu kódu?
- Jak zajistit **propustnost**?



# Původní vývojový model

- stabilní větev
- vývojová větev
- číslo verze: x.y.z (např. 2.4.14)
- stabilní: Even a newbie can use this.
- vývojová: Odd things may happen.



# Původní vývojový model

- stabilní větev
- vývojová větev
- číslo verze: x.y.z (např. 2.4.14)
  - stabilní: Even a newbie can use this.
  - vývojová: Odd things may happen.



# Původní vývojový model

- stabilní větev
- vývojová větev
- číslo verze: x.y.z (např. 2.4.14)
- stabilní: **Even** a newbie can use this.
- vývojová: Odd things may happen.



# Původní vývojový model

- stabilní větev
- vývojová větev
- číslo verze: x.y.z (např. 2.4.14)
- stabilní: Even a newbie can use this.
- vývojová: **Odd** things may happen.



# Release cycle

## Historická data

**2.2.0** – leden 1999

**2.4.0** – leden 2001

**2.6.0** – prosinec 2003

## Příliš dlouho!

- odradí uživatele (nové drivery)
- odradí vývojáře (jejich kód se nepoužívá)
- odradí distributory (musí používat část vývojové větve)

# Release cycle

## Historická data

**2.2.0** – leden 1999

**2.4.0** – leden 2001

**2.6.0** – prosinec 2003

## Příliš dlouho!

- odradí uživatele (nové drivery)
- odradí vývojáře (jejich kód se nepoužívá)
- odradí distributory (musí používat část vývojové větve)



# Release cycle

## Historická data

**2.2.0** – leden 1999

**2.4.0** – leden 2001

**2.6.0** – prosinec 2003

## Příliš dlouho!

- odradí uživatele (nové drivery)
- odradí vývojáře (jejich kód se nepoužívá)
- odradí distributory (musí používat část vývojové větve)





# Jak to řešit?

## Kratší vývojový cyklus!

- další vývoj ve „stabilní“ větvi 2.6
- větev 2.7 se neplánuje
- 2.6.n – 2.6.n+1: 2–3 měsíce

### Cíl:

Příliš se nevzdalovat od stabilního stavu.

# Jak to řešit?

## Kratší vývojový cyklus!

- další vývoj ve „stabilní“ větvi 2.6
- větev 2.7 se neplánuje
- 2.6.n – 2.6.n+1: 2–3 měsíce

### Cíl:

Příliš se nevzdalovat od stabilního stavu.

# Jak to řešit?

## Kratší vývojový cyklus!

- další vývoj ve „stabilní“ větvi 2.6
- větev 2.7 se neplánuje
- 2.6.n – 2.6.n+1: 2–3 měsíce

### Cíl:

Příliš se nevzdalovat od stabilního stavu.

# Jak to řešit?

## Kratší vývojový cyklus!

- další vývoj ve „stabilní“ větvi 2.6
- větev 2.7 se neplánuje
- 2.6.n – 2.6.n+1: 2–3 měsíce

### Cíl:

Příliš se nevzdalovat od stabilního stavu.

# Jak to řešit?

## Kratší vývojový cyklus!

- další vývoj ve „stabilní“ větvi 2.6
- větev 2.7 se neplánuje
- 2.6.n – 2.6.n+1: 2–3 měsíce

### Cíl:

Příliš se nevzdalovat od stabilního stavu.

# Časový rozvrh

- 2.6.n: otevře se merge window
- 2.6.n + 14 dní: 2.6.n+1-rc1 – jen opravy chyb
- 2.6.n+1-rcN: další stabilizace
- 2.6.n+1: další merge window



# Časový rozvrh

- 2.6.n: otevře se merge window
- 2.6.n + 14 dní: 2.6.n+1-rc1 – jen opravy chyb
- 2.6.n+1-rcN: další stabilizace
- 2.6.n+1: další merge window



# Časový rozvrh

- 2.6.n: otevře se merge window
- 2.6.n + 14 dní: 2.6.n+1-rc1 – jen opravy chyb
- 2.6.n+1-rcN: další stabilizace
- 2.6.n+1: další merge window





# Časový rozvrh

- 2.6.n: otevře se merge window
- 2.6.n + 14 dní: 2.6.n+1-rc1 – jen opravy chyb
- 2.6.n+1-rcN: další stabilizace
- 2.6.n+1: další merge window



# A co kritické chyby?



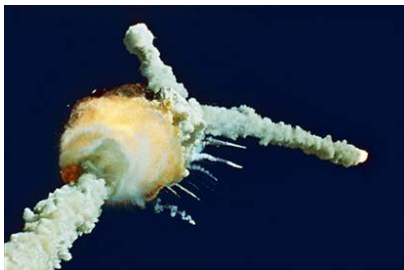
- 2.6.n.m
  - bezpečnostní problém
  - nebo chyba od 2.6.n-1
  - < 100 řádků patch

# A co kritické chyby?



- 2.6.n.m
- bezpečnostní problém
- nebo chyba od 2.6.n-1
- < 100 řádků patch

# A co kritické chyby?



- 2.6.n.m
- bezpečnostní problém
- nebo chyba od 2.6.n-1
- < 100 řádků patch

# A co kritické chyby?



- 2.6.n.m
- bezpečnostní problém
- nebo chyba od 2.6.n-1
- < 100 řádků patch

# Když dva měsíce nestačí

- Andrew Morton
- -mm tree
- „vývojová větev“
- pro dlouhodobější projekty



# Správa verzí

- až do verze 2.4: žádný verzovací systém
- Linux 2.5: BitKeeper

## BitKeeper

- closed source
  - jednotka práce: changeset
  - nelineární vývoj
  - slučování větví
- 
- 2.6.12 – ztráta licence BitKeeperu
  - Linus Torvalds píše verzovací systém: git

# Správa verzí

- až do verze 2.4: žádný verzovací systém
- Linux 2.5: BitKeeper

## BitKeeper

- closed source
  - jednotka práce: changeset
  - nelineární vývoj
  - slučování větví
- 
- 2.6.12 – ztráta licence BitKeeperu
  - Linus Torvalds píše verzovací systém: git



# Správa verzí

- až do verze 2.4: žádný verzovací systém
- Linux 2.5: BitKeeper

## BitKeeper

- closed source
  - jednotka práce: changeset
  - nelineární vývoj
  - slučování větví
- 
- 2.6.12 – ztráta licence BitKeeperu
  - Linus Torvalds píše verzovací systém: git

# Git



- content-addressable filesystem
- idea z VCS Monotone
- <http://git.or.cz/>
- rychlost!
- nelineární vývoj
- slučování větví
- kryptografická prokazatelnost historie

# Git



- content-addressable filesystem
- idea z VCS **Monotone**
- <http://git.or.cz/>
- rychlost!
- nelineární vývoj
- slučování větví
- kryptografická prokazatelnost historie

# Git



- content-addressable filesystem
- idea z VCS Monotone
- <http://git.or.cz/>
- rychlost!
- nelineární vývoj
- slučování větví
- kryptografická prokazatelnost historie

# Git



- content-addressable filesystem
- idea z VCS Monotone
- <http://git.or.cz/>
- rychlost!
- nelineární vývoj
- slučování větví
- kryptografická prokazatelnost historie

# Git



- content-addressable filesystem
- idea z VCS Monotone
- <http://git.or.cz/>
- rychlost!
- **nelineární** vývoj
- slučování větví
- kryptografická prokazatelnost historie

# Git



- content-addressable filesystem
- idea z VCS Monotone
- <http://git.or.cz/>
- rychlost!
- nelineární vývoj
- **slučování větví**
- kryptografická prokazatelnost historie

# Git



- content-addressable filesystem
- idea z VCS Monotone
- <http://git.or.cz/>
- rychlost!
- nelineární vývoj
- slučování větví
- kryptografická **prokazatelnost historie**



# Rychlost vývoje

Verze	M řádek	CSets	CSet/h	Vývojářů	Firem
2.6.11	6.6	4041	2.44	479	30
2.6.12	6.7	5565	2.15	704	38
2.6.13	6.9	4174	2.38	641	39
2.6.14	7.1	3931	2.69	632	45
2.6.15	7.2	5410	3.31	685	49
2.6.16	7.4	5734	3.10	782	56
2.6.17	7.5	6113	2.80	787	54
2.6.18	7.7	6791	2.98	904	60
2.6.19	7.9	7073	4.09	887	67
2.6.20	8.1	4983	3.05	730	75
2.6.21	8.2	5349	2.75	838	68

# Podíl firem na vývoji

Firma	Počet změn	Procenta
Nezjištěno	27976	47.3
Red Hat	6106	10.3
Novell	5923	10.0
Linux Foundation	4843	8.2
IBM	3991	6.7
Intel	2244	3.8
SGI	1353	2.3
NetApp	636	1.1
...	...	...

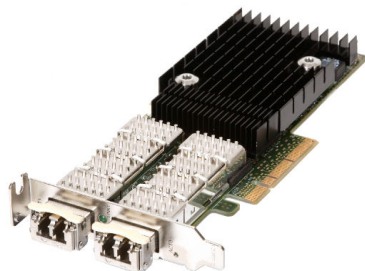
# Co je nového v kernelu?

1 Vývojový model Linuxu

**2 Co je nového v kernelu?**

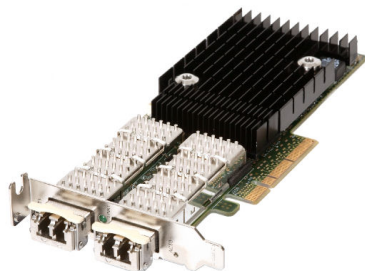
# Sít'ová rozhraní

- NAPI – nové rozhraní pro drivery
- zjednodušené zamykání
- IRQ mitigation, polling
- multi-queue NIC
- 2.6.23



# Sít'ová rozhraní

- NAPI – nové rozhraní pro drivery
- zjednodušené zamykání
- IRQ mitigation, polling
- multi-queue NIC
- 2.6.23



# Tickless kernel

- konfigurovatelné HZ
- nestačí, CPU se neuspí dostatečně
- tickless system
- není-li co dělat, spí se déle
- problém: desktopová prostředí („Halóóó, ještě nespíte?“)



# Tickless kernel

- konfigurovatelné HZ
- nestačí, CPU se neuspí dostatečně
- tickless system
- není-li co dělat, spí se déle
- problém: desktopová prostředí („Halóóó, ještě nespíte?“)



# Tickless kernel

- konfigurovatelné HZ
- nestačí, CPU se neuspí dostatečně
- tickless system
  - není-li co dělat, spí se déle
  - problém: desktopová prostředí („Halóóó, ještě nespíte?“)





# Tickless kernel

- konfigurovatelné HZ
- nestačí, CPU se neuspí dostatečně
- tickless system
- není-li co dělat, spí se déle
- problém: desktopová prostředí („Halóóó, ještě nespíte?“)



# Tickless kernel

- konfigurovatelné HZ
- nestačí, CPU se neuspí dostatečně
- tickless system
- není-li co dělat, spí se déle
- problém: desktopová prostředí („Halóóó, ještě nespíte?“)



# CPUfreq framework

- změna frekvence CPU za běhu
- **driver** – konkrétní CPU (nebo ACPI)
- **governor** – algoritmus výběru frekvence

## CPUfreq governors

- performance
- powersave
- userspace
- ondemand
- conservative



# CPUfreq framework

- změna frekvence CPU za běhu
- **driver** – konkrétní CPU (nebo ACPI)
- **governor** – algoritmus výběru frekvence

## CPUfreq governors

- performance
- powersave
- userspace
- ondemand
- conservative



# Inotify

- změna obsahu adresáře?
  - polling?
  - notifikace od kernelu: dnotify, SGI FAM.
  - problémy: sledování rozdílů, otevřené adresáře
  - řešení: inotify

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i umount notifikace!)
- pracuje i s podstromem

# Inotify

- změna obsahu adresáře
- polling?
- notifikace od kernelu: `dnotify`, SGI FAM.
- problémy: sledování rozdílů, otevřené adresáře
- řešení: `inotify`

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i `umount` notifikace!)
- pracuje i s podstromem

# Inotify

- změna obsahu adresáře
- polling? **fuj!**
- notifikace od kernelu: dnotify, SGI FAM.
- problémy: sledování rozdílů, otevřené adresáře
- řešení: inotify

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i umount notifikace!)
- pracuje i s podstromem

# Inotify

- změna obsahu adresáře
- polling? fuj!
- notifikace od kernelu: `dnotify`, SGI FAM.
- problémy: sledování rozdílů, otevřené adresáře
- řešení: `inotify`

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i `umount` notifikace!)
- pracuje i s podstromem



# Inotify

- změna obsahu adresáře
- polling? fuj!
- notifikace od kernelu: `dnotify`, SGI FAM.
- problémy: sledování rozdílů, otevřené adresáře
- řešení: `inotify`

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i `umount` notifikace!)
- pracuje i s podstromem

# Inotify

- změna obsahu adresáře
- polling? fuj!
- notifikace od kernelu: `dnotify`, SGI FAM.
- problémy: sledování rozdílů, otevřené adresáře
- řešení: `inotify`

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i `umount` notifikace!)
- pracuje i s podstromem

# Inotify

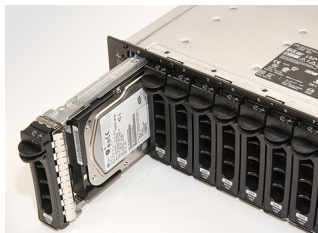
- změna obsahu adresáře
- polling? fuj!
- notifikace od kernelu: `dnotify`, SGI FAM.
- problémy: sledování rozdílů, otevřené adresáře
- řešení: `inotify`

## Vlastnosti inotify

- Linux 2.6.13+
- notifikace čtením z deskriptoru
- nezamyká adresář (i `umount` notifikace!)
- pracuje i s podstromem

# Nové souborové systémy

- OCFS2 – clusterový FS
- GFS – clusterový FS
- AFS
- Reiser4 – pluginy, více proudů, ...
- FUSE – v user-space
- Ext4 – vývoj Ext2/Ext3 pokračuje
- ...



# Virtualizace

## XEN

- paravirtualizace
- není plně v jádře
- nejpokročilejší

## KVM

- plná virtualizace
- vyžaduje HW podporu
- emulátor HW, např.  
Qemu

## Iguest

- vyslov „Rustyvisor“
- paravirtualizace
- nejjednodušší



# Virtualizace

## XEN

- paravirtualizace
- není plně v jádře
- nejpokročilejší

## KVM

- plná virtualizace
- vyžaduje HW podporu
- emulátor HW, např. Qemu

## Iguest

- vyslov „Rustyvisor“
- paravirtualizace
- nejjednodušší



# Virtualizace

## XEN

- paravirtualizace
- není plně v jádře
- nejpokročilejší

## KVM

- plná virtualizace
- vyžaduje HW podporu
- emulátor HW, např.  
Qemu

## Iguest

- vyslov „Rustyvisor“
- paravirtualizace
- nejjednodušší



# Shrnutí

- Jádru 2.6 má nový vývojový model.
- Vývoj je fakt rychlý.
- Verze 2.7 se zatím neplánuje.



# Shrnutí

- Jádro 2.6 má nový vývojový model.
- Vývoj je **fakt** rychlý.
- Verze 2.7 se zatím neplánuje.

# Shrnutí

- Jádro 2.6 má nový vývojový model.
- Vývoj je fakt rychlý.
- Verze 2.7 se zatím neplánuje.

Děkuji za pozornost.