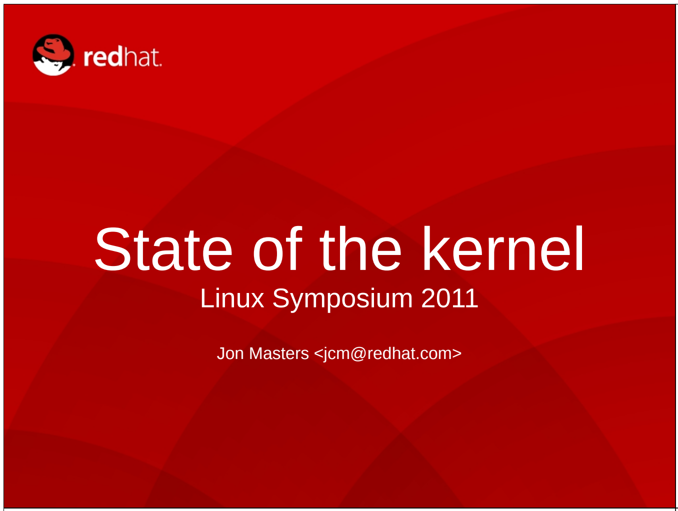
State of the kernel

John C. Masters
Red Hat Inc.
jcm@redhat.com

Abstract

Slides from the talk follow.

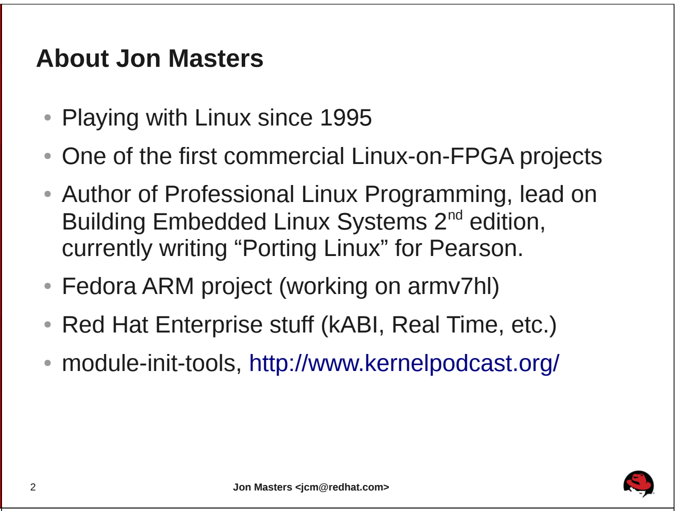


redhat.

State of the kernel

Linux Symposium 2011

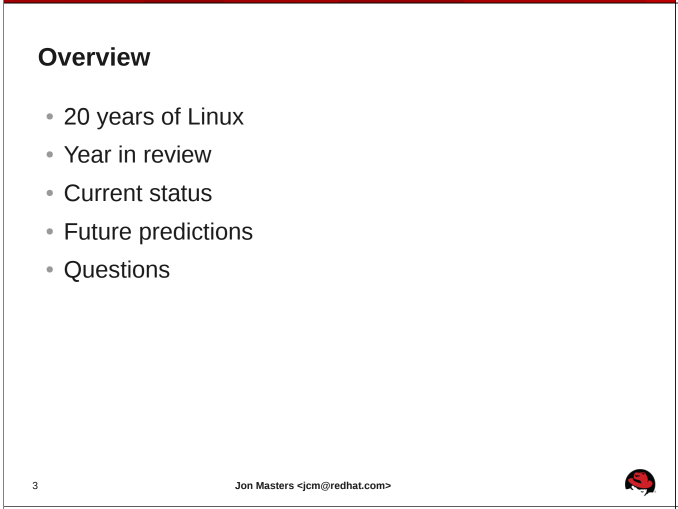
Jon Masters <jcm@redhat.com>



About Jon Masters

- Playing with Linux since 1995
- One of the first commercial Linux-on-FPGA projects
- Author of Professional Linux Programming, lead on Building Embedded Linux Systems 2nd edition, currently writing “Porting Linux” for Pearson.
- Fedora ARM project (working on armv7hl)
- Red Hat Enterprise stuff (kABI, Real Time, etc.)
- module-init-tools, http://www.kernelpodcast.org/

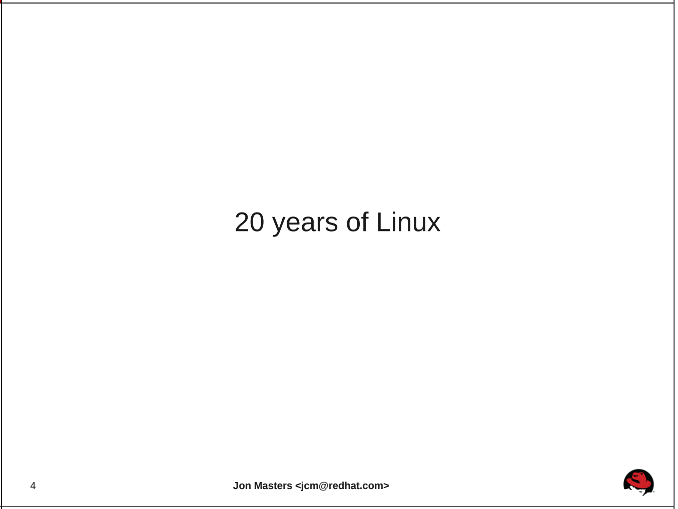
2 Jon Masters <jcm@redhat.com>



Overview

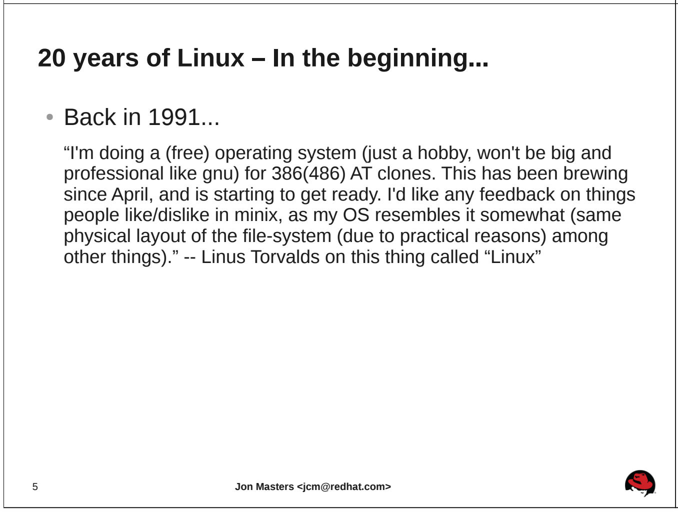
- 20 years of Linux
- Year in review
- Current status
- Future predictions
- Questions

3 Jon Masters <jcm@redhat.com>



20 years of Linux

4 Jon Masters <jcm@redhat.com>

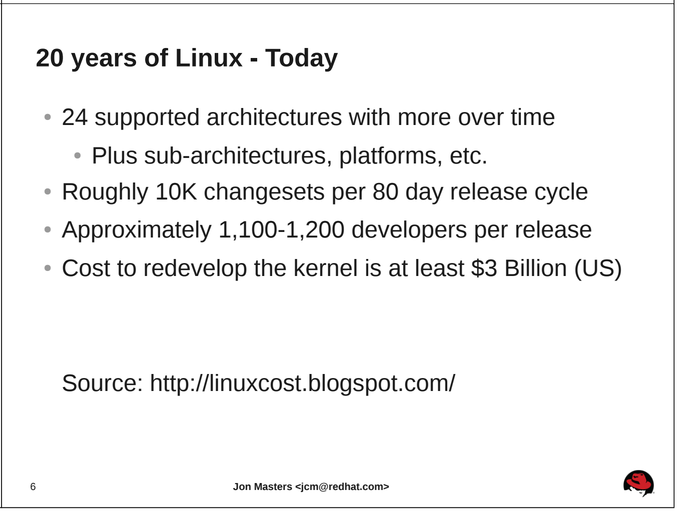


20 years of Linux – In the beginning...

- Back in 1991...

  “I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones. This has been brewing since April, and is starting to get ready. I'd like any feedback on things people like/dislike in minix, as my OS resembles it somewhat (same physical layout of the file-system (due to practical reasons) among other things).” -- Linus Torvalds on this thing called “Linux”

5 Jon Masters <jcm@redhat.com>

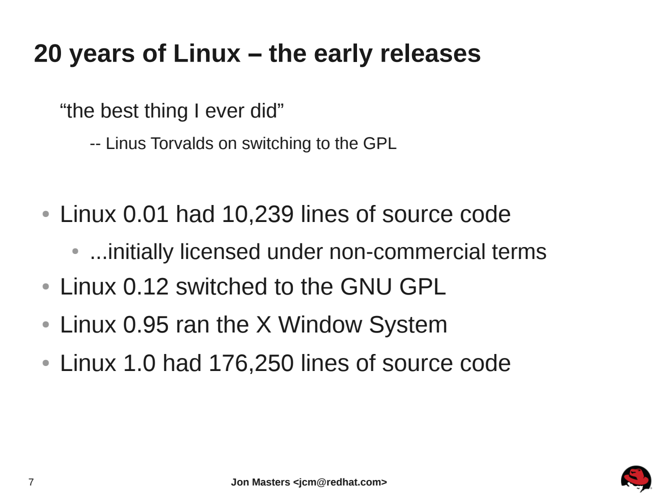


20 years of Linux - Today

- 24 supported architectures with more over time
  - Plus sub-architectures, platforms, etc.
- Roughly 10K changesets per 80 day release cycle
- Approximately 1,100-1,200 developers per release
- Cost to redevelop the kernel is at least $3 Billion (US)

Source: http://linuxcost.blogspot.com/

6 Jon Masters <jcm@redhat.com>



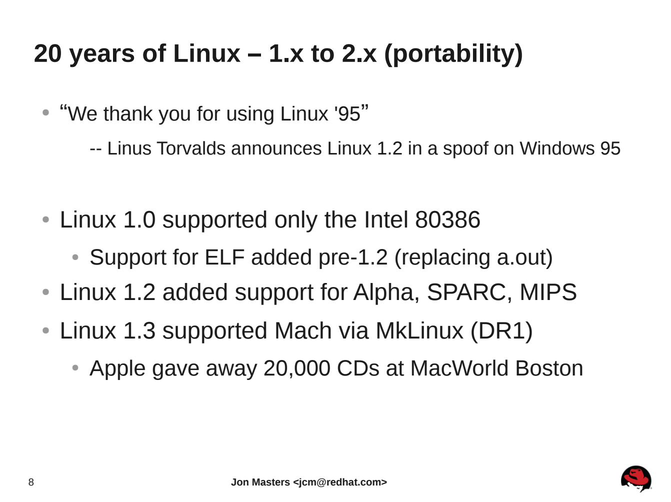
20 years of Linux – the early releases

"the best thing I ever did"

-- Linus Torvalds on switching to the GPL

- Linux 0.01 had 10,239 lines of source code
  - ...initially licensed under non-commercial terms
- Linux 0.12 switched to the GNU GPL
- Linux 0.95 ran the X Window System
- Linux 1.0 had 176,250 lines of source code

7 Jon Masters <jcm@redhat.com>



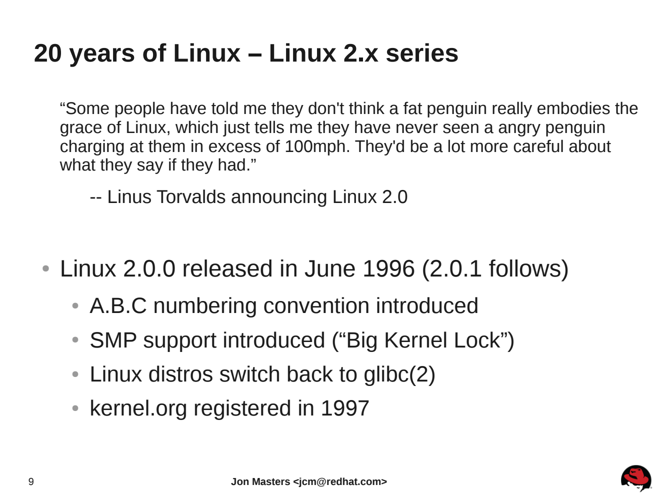
20 years of Linux – 1.x to 2.x (portability)

- "We thank you for using Linux '95"

  -- Linus Torvalds announces Linux 1.2 in a spoof on Windows 95

- Linux 1.0 supported only the Intel 80386
  - Support for ELF added pre-1.2 (replacing a.out)
- Linux 1.2 added support for Alpha, SPARC, MIPS
- Linux 1.3 supported Mach via MkLinux (DR1)
  - Apple gave away 20,000 CDs at MacWorld Boston

8 Jon Masters <jcm@redhat.com>



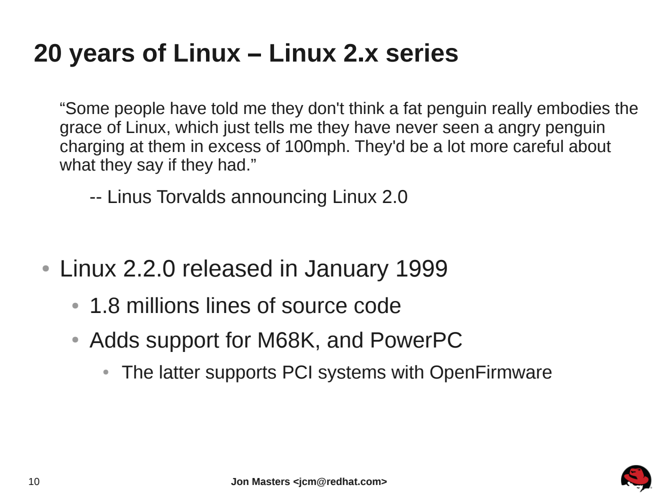
20 years of Linux – Linux 2.x series

"Some people have told me they don't think a fat penguin really embodies the grace of Linux, which just tells me they have never seen a angry penguin charging at them in excess of 100mph. They'd be a lot more careful about what they say if they had."

-- Linus Torvalds announcing Linux 2.0

- Linux 2.0.0 released in June 1996 (2.0.1 follows)
  - A.B.C numbering convention introduced
  - SMP support introduced ("Big Kernel Lock")
  - Linux distros switch back to glibc(2)
  - kernel.org registered in 1997

9 Jon Masters <jcm@redhat.com>



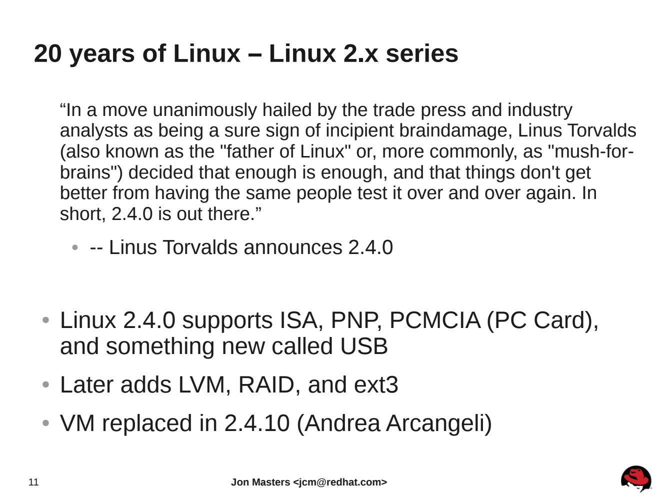
20 years of Linux – Linux 2.x series

"Some people have told me they don't think a fat penguin really embodies the grace of Linux, which just tells me they have never seen a angry penguin charging at them in excess of 100mph. They'd be a lot more careful about what they say if they had."

-- Linus Torvalds announcing Linux 2.0

- Linux 2.2.0 released in January 1999
  - 1.8 millions lines of source code
  - Adds support for M68K, and PowerPC
    - The latter supports PCI systems with OpenFirmware

10 Jon Masters <jcm@redhat.com>



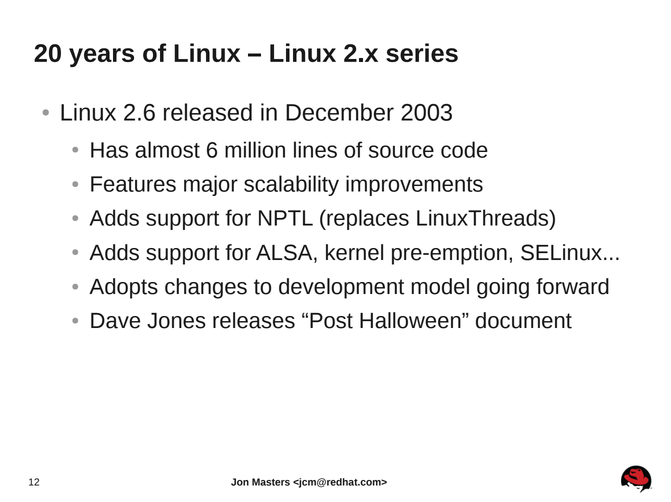
20 years of Linux – Linux 2.x series

"In a move unanimously hailed by the trade press and industry analysts as being a sure sign of incipient braindamage, Linus Torvalds (also known as the "father of Linux" or, more commonly, as "mush-for-brains") decided that enough is enough, and that things don't get better from having the same people test it over and over again. In short, 2.4.0 is out there."

- -- Linus Torvalds announces 2.4.0

- Linux 2.4.0 supports ISA, PNP, PCMCIA (PC Card), and something new called USB
- Later adds LVM, RAID, and ext3
- VM replaced in 2.4.10 (Andrea Arcangeli)

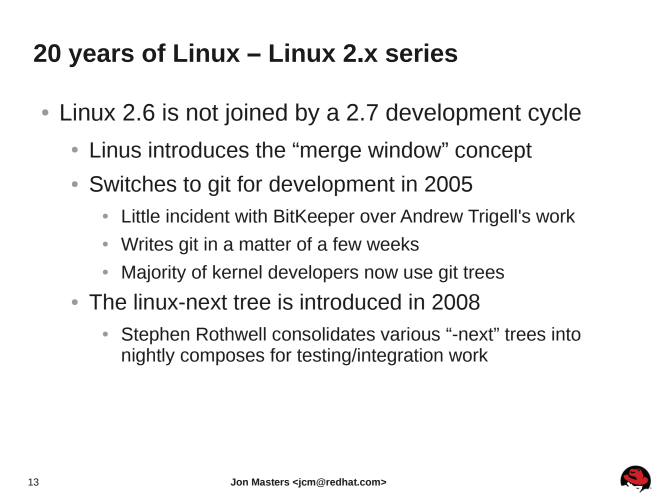
11 Jon Masters <jcm@redhat.com>



20 years of Linux – Linux 2.x series

- Linux 2.6 released in December 2003
  - Has almost 6 million lines of source code
  - Features major scalability improvements
  - Adds support for NPTL (replaces LinuxThreads)
  - Adds support for ALSA, kernel pre-emption, SELinux...
  - Adopts changes to development model going forward
  - Dave Jones releases "Post Halloween" document

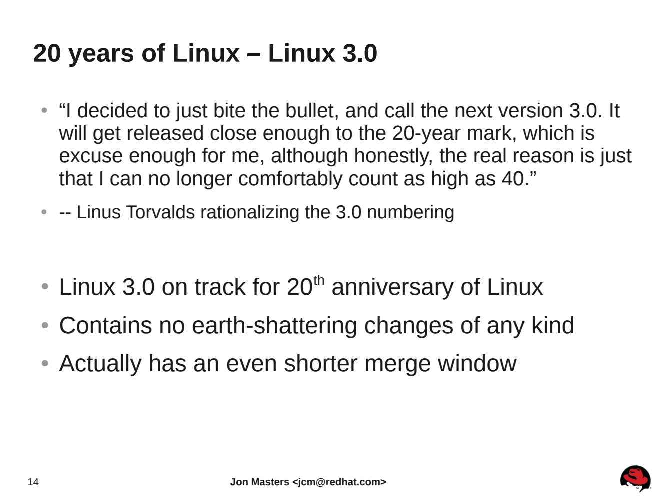
12 Jon Masters <jcm@redhat.com>



20 years of Linux – Linux 2.x series

- Linux 2.6 is not joined by a 2.7 development cycle
  - Linus introduces the “merge window” concept
  - Switches to git for development in 2005
    - Little incident with BitKeeper over Andrew Trigell's work
    - Writes git in a matter of a few weeks
    - Majority of kernel developers now use git trees
  - The linux-next tree is introduced in 2008
    - Stephen Rothwell consolidates various “-next” trees into nightly composes for testing/integration work

13 Jon Masters <jcm@redhat.com>



20 years of Linux – Linux 3.0

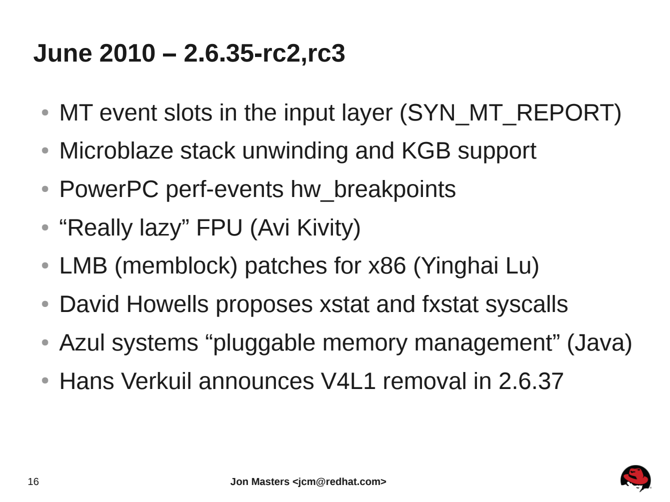
- “I decided to just bite the bullet, and call the next version 3.0. It will get released close enough to the 20-year mark, which is excuse enough for me, although honestly, the real reason is just that I can no longer comfortably count as high as 40.”
- -- Linus Torvalds rationalizing the 3.0 numbering

- Linux 3.0 on track for 20th anniversary of Linux
- Contains no earth-shattering changes of any kind
- Actually has an even shorter merge window

14 Jon Masters <jcm@redhat.com>

Year in Review

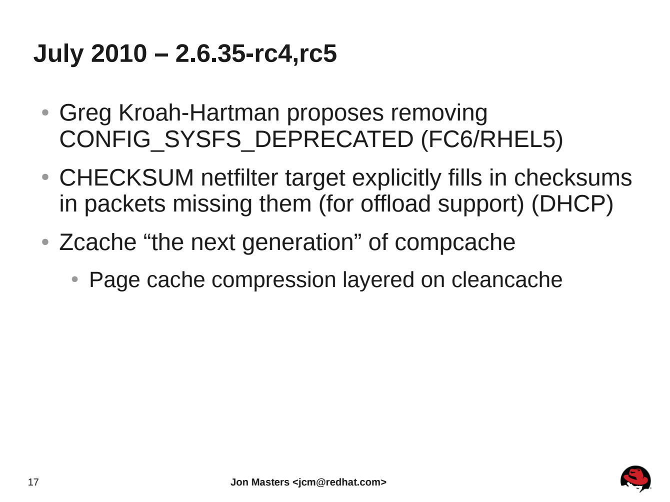
15 Jon Masters <jcm@redhat.com>



June 2010 – 2.6.35-rc2,rc3

- MT event slots in the input layer (SYN_MT_REPORT)
- Microblaze stack unwinding and KGB support
- PowerPC perf-events hw_breakpoints
- “Really lazy” FPU (Avi Kivity)
- LMB (memblock) patches for x86 (Yinghai Lu)
- David Howells proposes xstat and fxstat syscalls
- Azul systems “pluggable memory management” (Java)
- Hans Verkuil announces V4L1 removal in 2.6.37

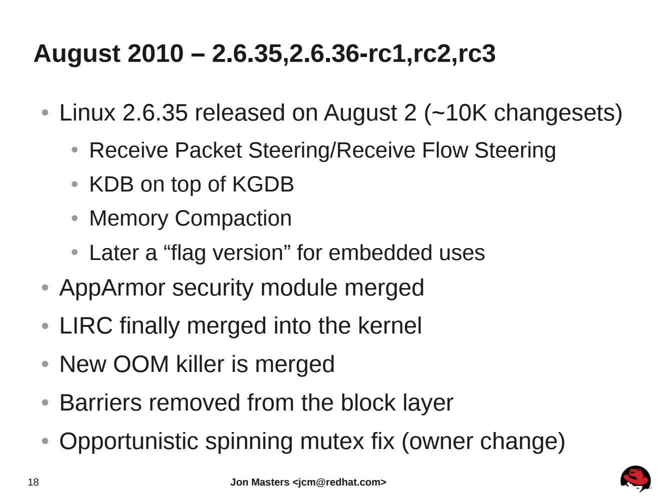
16 Jon Masters <jcm@redhat.com>



July 2010 – 2.6.35-rc4,rc5

- Greg Kroah-Hartman proposes removing CONFIG_SYSFS_DEPRECATED (FC6/RHEL5)
- CHECKSUM netfilter target explicitly fills in checksums in packets missing them (for offload support) (DHCP)
- Zcache “the next generation” of compcache
  - Page cache compression layered on cleancache

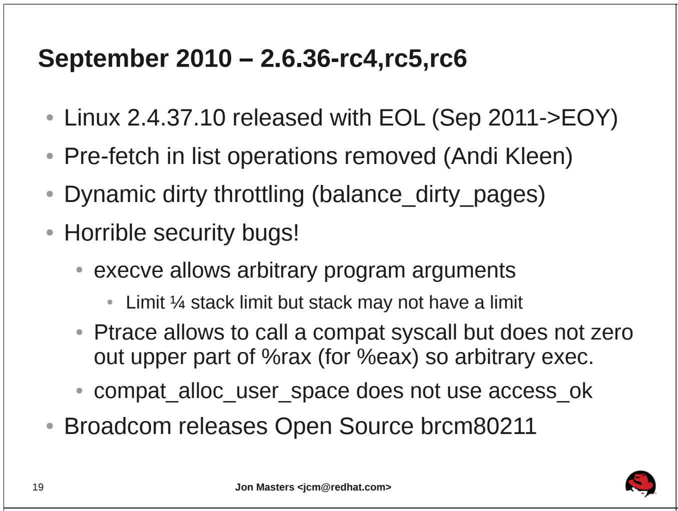
17 Jon Masters <jcm@redhat.com>



August 2010 – 2.6.35,2.6.36-rc1,rc2,rc3

- Linux 2.6.35 released on August 2 (~10K changesets)
  - Receive Packet Steering/Receive Flow Steering
  - KDB on top of KGDB
  - Memory Compaction
  - Later a “flag version” for embedded uses
- AppArmor security module merged
- LIRC finally merged into the kernel
- New OOM killer is merged
- Barriers removed from the block layer
- Opportunistic spinning mutex fix (owner change)

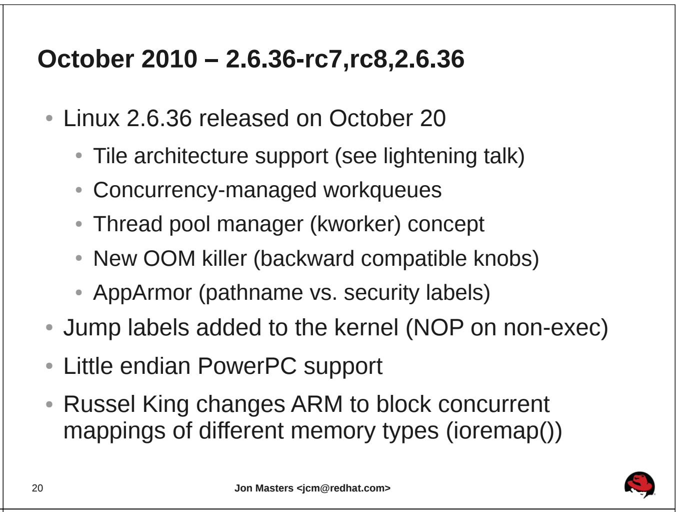
18 Jon Masters <jcm@redhat.com>



September 2010 – 2.6.36-rc4,rc5,rc6

- Linux 2.4.37.10 released with EOL (Sep 2011->EOY)
- Pre-fetch in list operations removed (Andi Kleen)
- Dynamic dirty throttling (balance_dirty_pages)
- Horrible security bugs!
  - execve allows arbitrary program arguments
    - Limit ¼ stack limit but stack may not have a limit
  - Ptrace allows to call a compat syscall but does not zero out upper part of %rax (for %eax) so arbitrary exec.
  - compat_alloc_user_space does not use access_ok
- Broadcom releases Open Source brcm80211

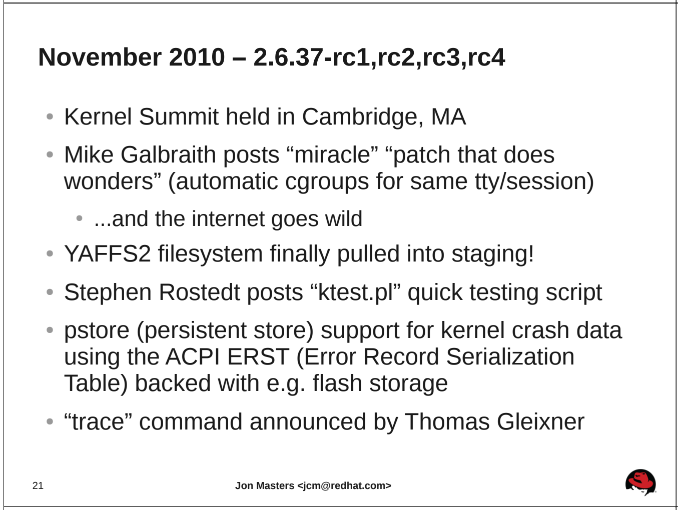
19 Jon Masters <jcm@redhat.com>



October 2010 – 2.6.36-rc7,rc8,2.6.36

- Linux 2.6.36 released on October 20
  - Tile architecture support (see lightening talk)
  - Concurrency-managed workqueues
  - Thread pool manager (kworker) concept
  - New OOM killer (backward compatible knobs)
  - AppArmor (pathname vs. security labels)
- Jump labels added to the kernel (NOP on non-exec)
- Little endian PowerPC support
- Russel King changes ARM to block concurrent mappings of different memory types (ioremap())

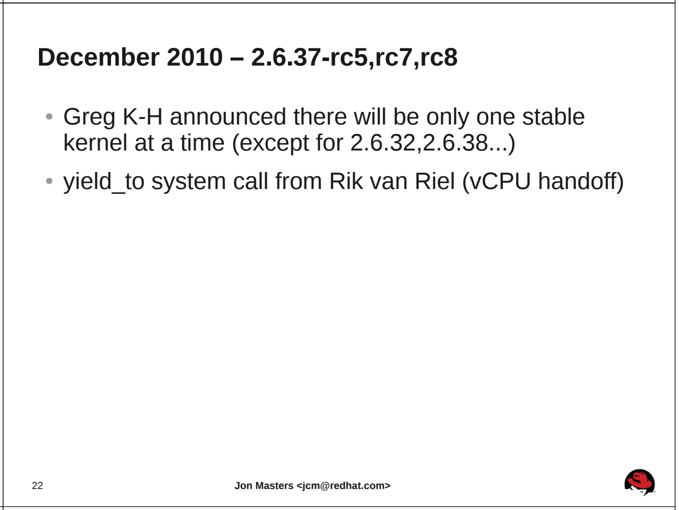
20 Jon Masters <jcm@redhat.com>



November 2010 – 2.6.37-rc1,rc2,rc3,rc4

- Kernel Summit held in Cambridge, MA
- Mike Galbraith posts “miracle” “patch that does wonders” (automatic cgroups for same tty/session)
  - ...and the internet goes wild
- YAFFS2 filesystem finally pulled into staging!
- Stephen Rostedt posts “ktest.pl” quick testing script
- pstore (persistent store) support for kernel crash data using the ACPI ERST (Error Record Serialization Table) backed with e.g. flash storage
- “trace” command announced by Thomas Gleixner

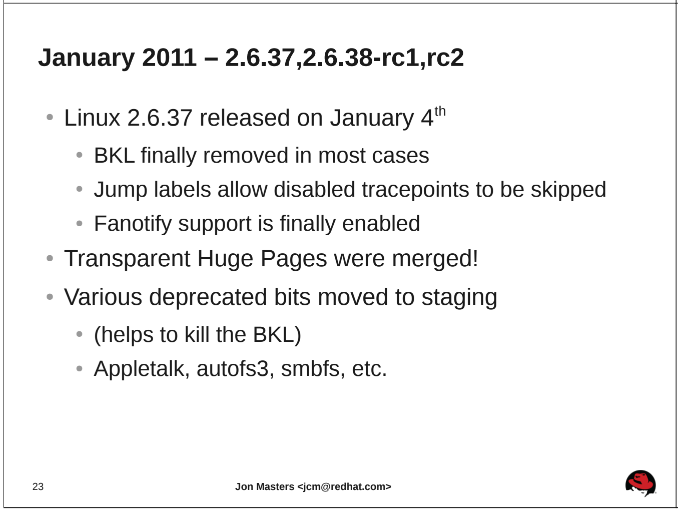
21 Jon Masters <jcm@redhat.com>



December 2010 – 2.6.37-rc5,rc7,rc8

- Greg K-H announced there will be only one stable kernel at a time (except for 2.6.32,2.6.38...)
- yield_to system call from Rik van Riel (vCPU handoff)

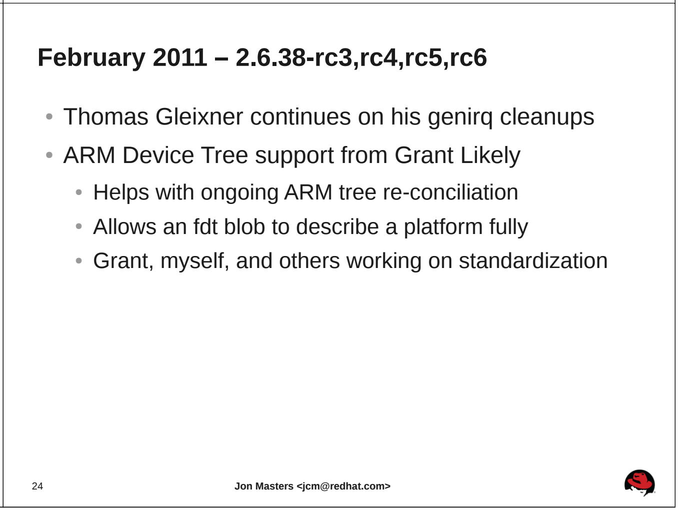
22 Jon Masters <jcm@redhat.com>



January 2011 – 2.6.37,2.6.38-rc1,rc2

- Linux 2.6.37 released on January 4th
  - BKL finally removed in most cases
  - Jump labels allow disabled tracepoints to be skipped
  - Fanotify support is finally enabled
- Transparent Huge Pages were merged!
- Various deprecated bits moved to staging
  - (helps to kill the BKL)
  - Appletalk, autofs3, smbfs, etc.

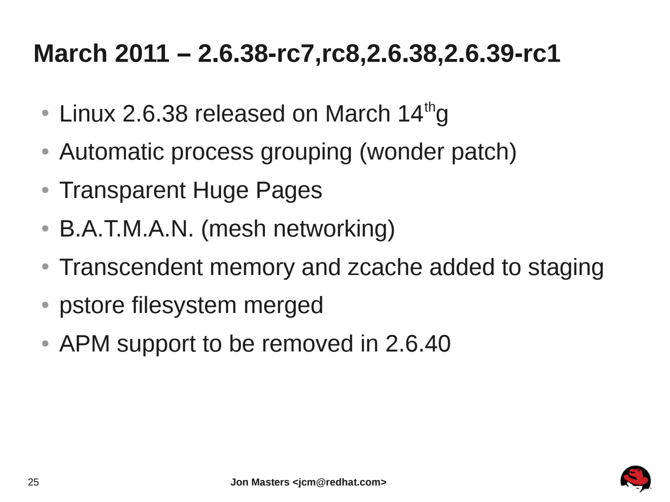
23 Jon Masters <jcm@redhat.com>



February 2011 – 2.6.38-rc3,rc4,rc5,rc6

- Thomas Gleixner continues on his genirq cleanups
- ARM Device Tree support from Grant Likely
  - Helps with ongoing ARM tree re-conciliation
  - Allows an fdt blob to describe a platform fully
  - Grant, myself, and others working on standardization

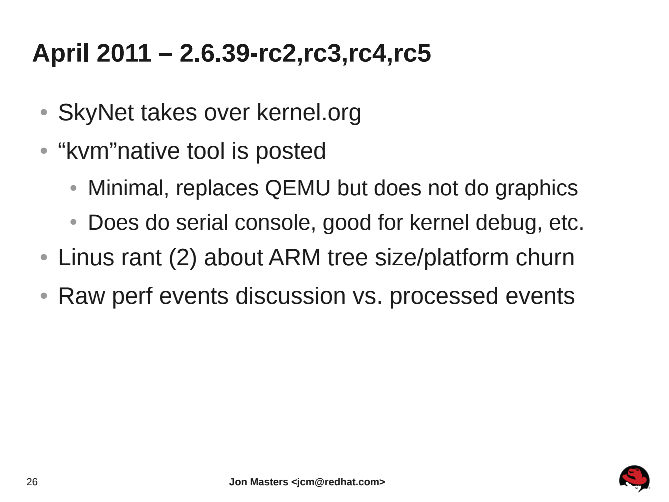
24 Jon Masters <jcm@redhat.com>



March 2011 – 2.6.38-rc7,rc8,2.6.38,2.6.39-rc1

- Linux 2.6.38 released on March 14thg
- Automatic process grouping (wonder patch)
- Transparent Huge Pages
- B.A.T.M.A.N. (mesh networking)
- Transcendent memory and zcache added to staging
- pstore filesystem merged
- APM support to be removed in 2.6.40

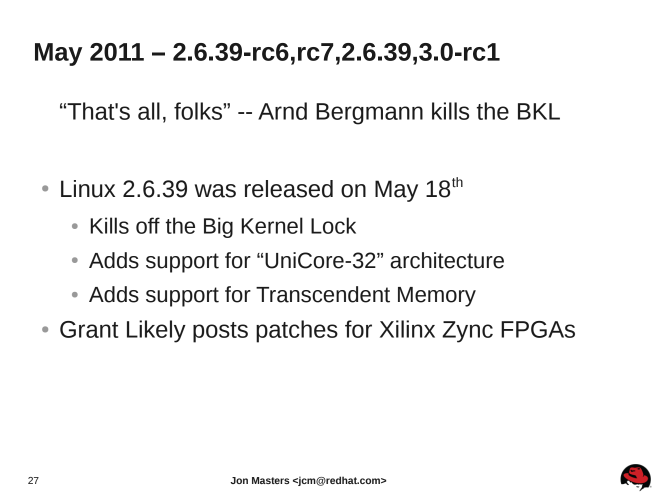
25 Jon Masters <jcm@redhat.com>



April 2011 – 2.6.39-rc2,rc3,rc4,rc5

- SkyNet takes over kernel.org
- “kvm”native tool is posted
  - Minimal, replaces QEMU but does not do graphics
  - Does do serial console, good for kernel debug, etc.
- Linus rant (2) about ARM tree size/platform churn
- Raw perf events discussion vs. processed events

26 Jon Masters <jcm@redhat.com>

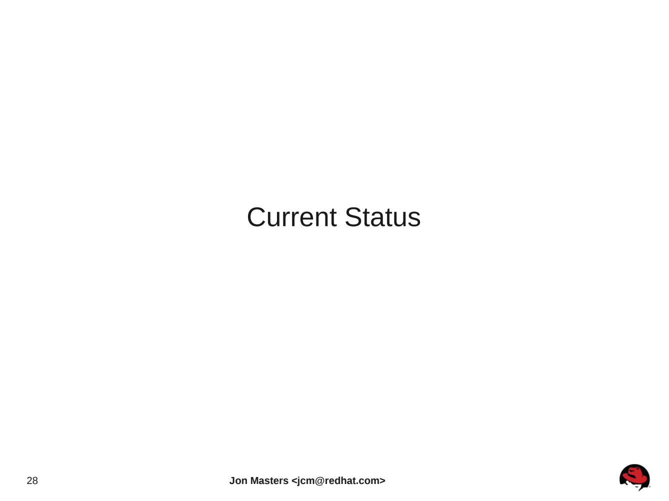


May 2011 – 2.6.39-rc6,rc7,2.6.39,3.0-rc1

“That's all, folks” -- Arnd Bergmann kills the BKL

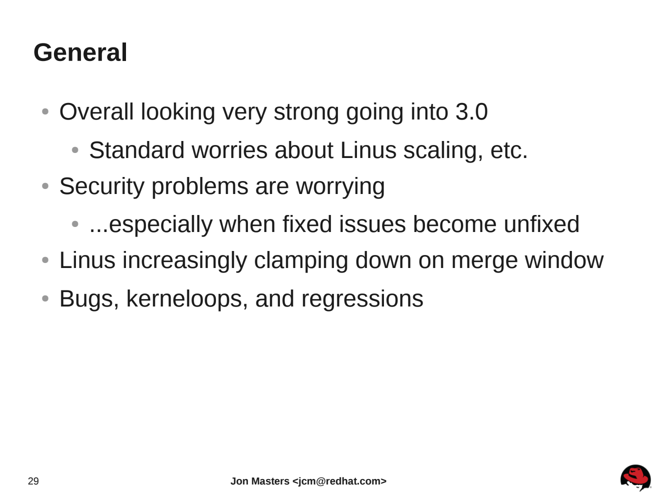
- Linux 2.6.39 was released on May 18th
  - Kills off the Big Kernel Lock
  - Adds support for “UniCore-32” architecture
  - Adds support for Transcendent Memory
- Grant Likely posts patches for Xilinx Zync FPGAs

27 Jon Masters <jcm@redhat.com>



Current Status

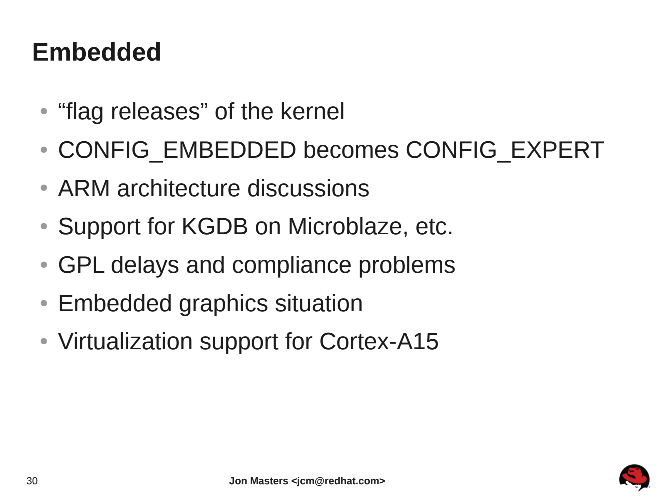
28 Jon Masters <jcm@redhat.com>



General

- Overall looking very strong going into 3.0
  - Standard worries about Linus scaling, etc.
- Security problems are worrying
  - ...especially when fixed issues become unfixed
- Linus increasingly clamping down on merge window
- Bugs, kerneloops, and regressions

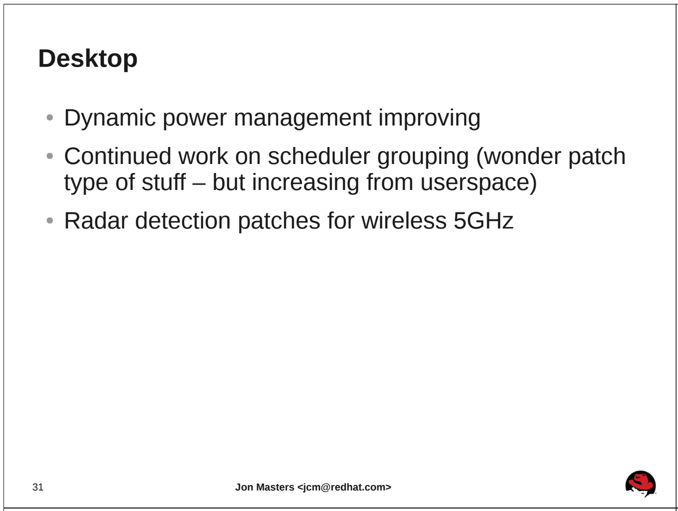
29 Jon Masters <jcm@redhat.com>



Embedded

- “flag releases” of the kernel
- CONFIG_EMBEDDED becomes CONFIG_EXPERT
- ARM architecture discussions
- Support for KGDB on Microblaze, etc.
- GPL delays and compliance problems
- Embedded graphics situation
- Virtualization support for Cortex-A15

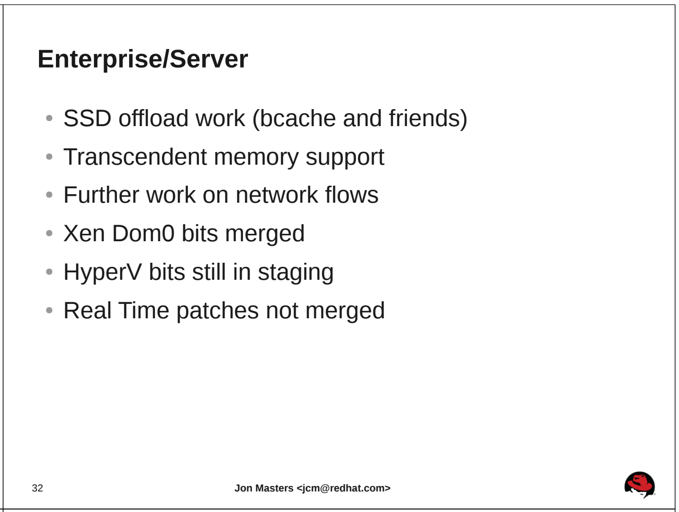
30 Jon Masters <jcm@redhat.com>



Desktop

- Dynamic power management improving
- Continued work on scheduler grouping (wonder patch type of stuff – but increasing from userspace)
- Radar detection patches for wireless 5GHz

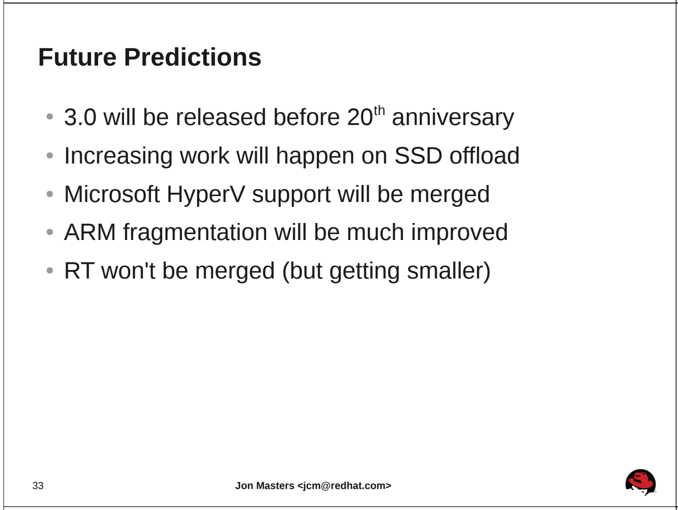
31 Jon Masters <jcm@redhat.com>



Enterprise/Server

- SSD offload work (bcache and friends)
- Transcendent memory support
- Further work on network flows
- Xen Dom0 bits merged
- HyperV bits still in staging
- Real Time patches not merged

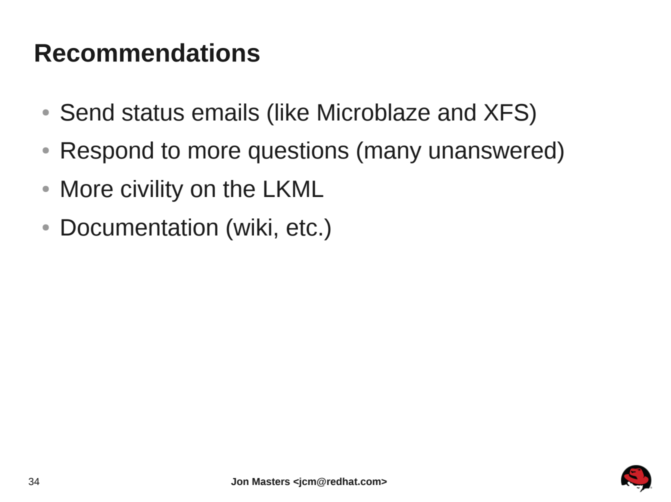
32 Jon Masters <jcm@redhat.com>



Future Predictions

- 3.0 will be released before 20th anniversary
- Increasing work will happen on SSD offload
- Microsoft HyperV support will be merged
- ARM fragmentation will be much improved
- RT won't be merged (but getting smaller)

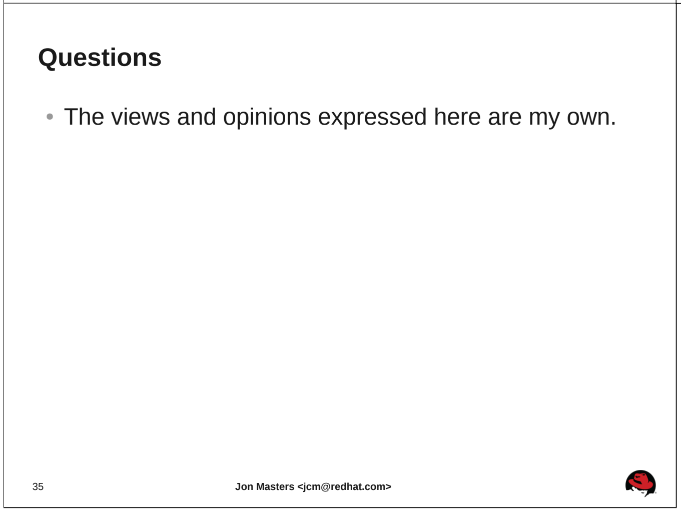
33 Jon Masters <jcm@redhat.com>



Recommendations

- Send status emails (like Microblaze and XFS)
- Respond to more questions (many unanswered)
- More civility on the LKML
- Documentation (wiki, etc.)

34 Jon Masters <jcm@redhat.com>



Questions

- The views and opinions expressed here are my own.

35 Jon Masters <jcm@redhat.com>