

# Speed up boot through disk sorting

## **Workout**

FOSS.IN 2008

Nikanth Karthikesan

Linux Kernel Engineer, Suse Labs, Novell

**Novell.**<sup>®</sup>

# Agenda

- Introduction
  - Basic Idea
  - Relocate the blocks
  - Device Mapper – the virtualization layer of disk
  - How they compare
- Discussion:
  - Practical Use cases
  - Best/right approach
- Workout
  - Current state
  - Work on TODOs

# Basic Idea

- Both relative and absolute position of data on disk influences performance
- Improve performance by relocating disk blocks



# Relocate the blocks

- Fcache
  - blktrace
- FS specific defragmentation tools
- E2remapblocks
  - E2block2file & blktrace

# Using Device Mapper

# Using Device Mapper

- Use blktrace to trace the accesses
- Compute the ideal layout
- Sort the blocks in the disk, in the ideal layout
- Setup DM table

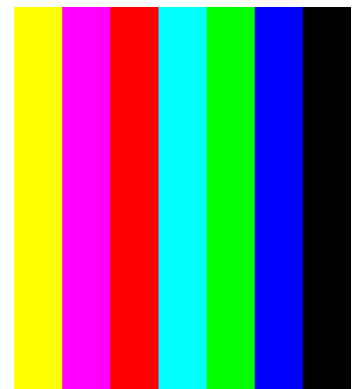
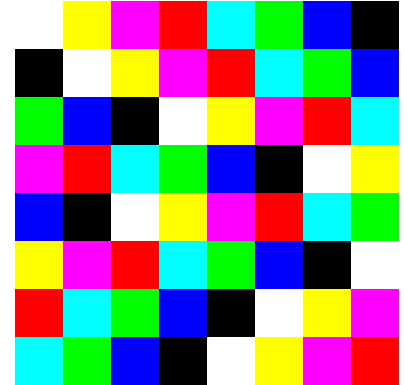
# Blktrace

- Blktrace
  - allows tracing IO requests going to disk
  - Blkparse converts to user readable format
  - `blktrace /dev/sda -o - -a READ -a WRITE | blkparse - -f "%M: %m %d %a %C %S+%n\n"`



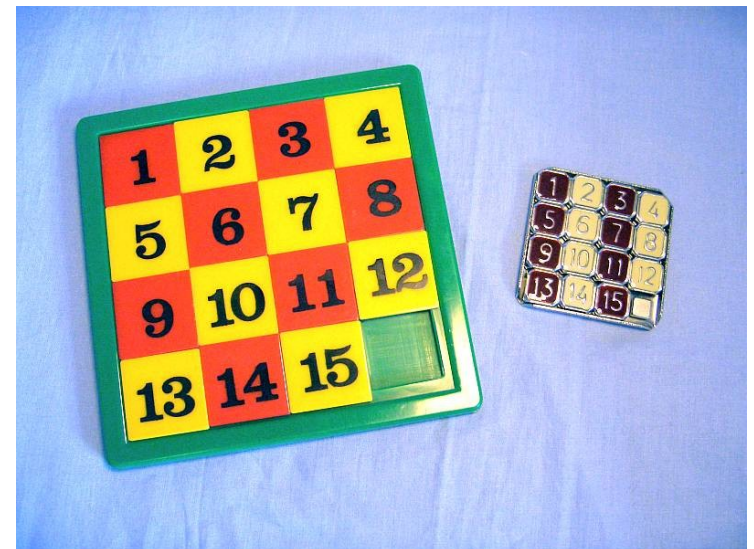
# Ideal disk layout

- Compute the ideal layout
  - Set of utilities to extract the
    - > Consecutive sectors accessed
    - > Sector pairs for seeks performed
  - And compute an ideal layout
    - > Do not break existing consecutive sectors
    - > Rearrange to reduce seeks
    - > If data in the areas of disk with lesser throughput is used more often than on the better location, remap it



# Setup DM

- Modify disk
  - Move disk blocks around so that they are in the ideal layout calculated
  - Setup dm table so that the fs layer above does not know about this change below its feet



# How it compares

(+) Filesystem independent

(-) But filesystem might try to do something and result in more chaos

(+) Based on actual(historic) data

(+) Data from FIEMAP/FIBMAP ioctl can also be used instead of/in addition to blktrace

(+) Tuned for that particular system – memory size, cache available to the process, etc...

(+) No separate Partition like Fcache

(-) Optimizing for one application might hinder other applications

(-) offline

(+) Filesystems do it online and hence the need for this ;-)

The background of the slide is a solid blue color with a pattern of diagonal lines in various shades of blue, creating a sense of motion and depth. The lines are more densely packed on the right side and become more sparse towards the left.

Questions

# Discuss

- Any practical use cases?
  - Read-only media? WORM, write in place? Some specific application that bangs the disk the same way again and again?
  - When a large application is started, the icons, libraries, binary stay in different directories, hence start up generates the same sequence of reads
  - Applications can be fixed/improved, but there are many.

# Discuss

- Right, filesystem should do this
  - but there is currently no way to tell a filesystem that these would be accessed sequentially, or so
  - Also they normally do not try to move blocks when off-line, so cannot do massive sorting this can do – defragmentors can
- Just a way to override filesystem's allocation strategy
- But wait SSDs are around the corner, who cares about seeks?!



Workout

# Workout

- <http://gitorious.org/projects/disk-sort>
- Current state – just a prototype
  - Couple of programs
    - > to convert the blktrace output to an ideal layout
    - > to shuffle the blocks in the disk to this new “ideal” layout
  - Does not consider seek distance into account! :(
  - Does not optimize for minimal disk shuffling! :(
  - Ugly code – not enough eyeballs
- Try optimizing the disk for specific workloads
  - Brave enough to loose data, better try on spare partitions/disks

# TODO

- Knowledge Req'd: Algorithms and data structures, C, Linux system calls and understanding of the problem ;)
  - Optimize/change the algorithms used
  - Change the programs or start from scratch to do it faster and better [currently uses sqlite – but would be better to do it in memory]
  - Add support for considering seek distance into account
  - Minimize the disk movement required
- Knowledge Req'd: Device mapper, lvm
  - Integrate with lvm2 (<ftp://sources.redhat.com/pub/lvm2/>)
    - > Req'd for boot optimization
  - Write a easy-to-use wrapper to automate everything
- Better undo support
- And whatever that came out of the discussion

**Novell®**

## **Unpublished Work of Novell, Inc. All Rights Reserved.**

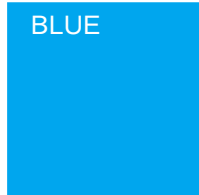
This work is an unpublished work and contains confidential, proprietary, and trade secret information of Novell, Inc. Access to this work is restricted to Novell employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of Novell, Inc. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

## **General Disclaimer**

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. Novell, Inc. makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for Novell products remains at the sole discretion of Novell. Further, Novell, Inc. reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All Novell marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.



# Color Palette



BLUE

RGB  
0 166 238



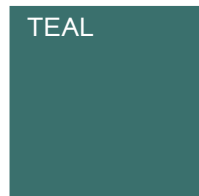
RED

RGB  
224 0 0



ORANGE

RGB  
230 120 20



TEAL

RGB  
50 118 109



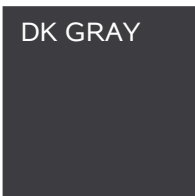
GREEN

RGB  
98 158 31



YELLOW

RGB  
255 221 0



DK GRAY

RGB  
60 60 65



MD GRAY

RGB  
90 90 100



LT GRAY

RGB  
204 204 205

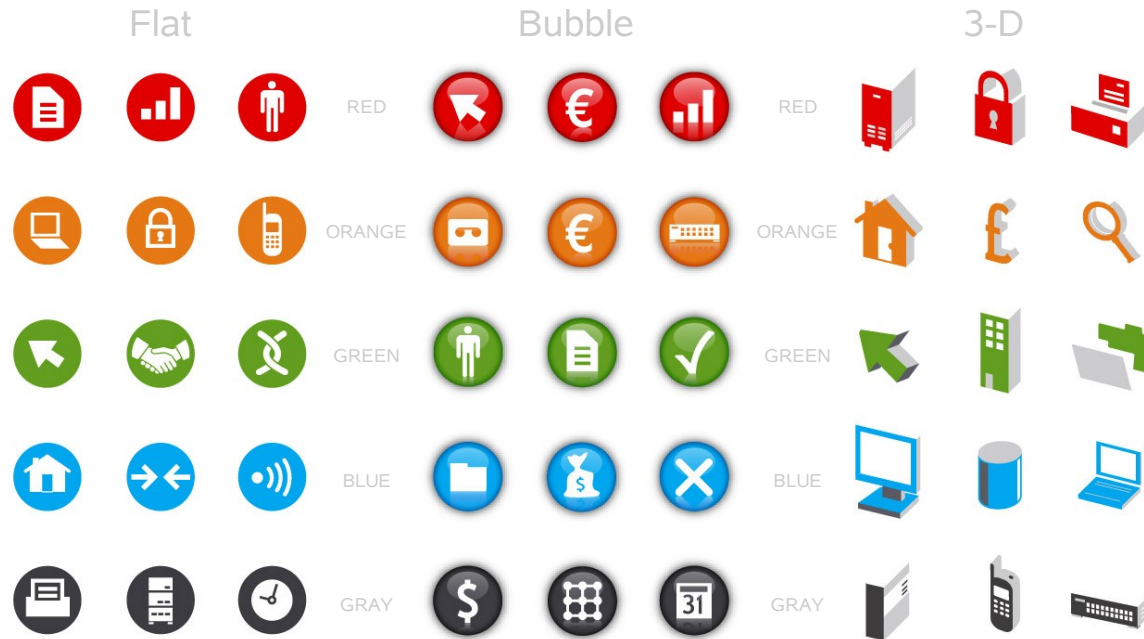
## Note:

The gray dotted-line box represents the margins or “working area” into which all text and most graphics and diagrams should conform.

## How to Add Novell Colors to Your OpenOffice Color Palette:

1. Go to the “Tools” menu
2. Select “Options”
3. Expand “OpenOffice.org”
4. Select “Colors”
5. Delete existing colors (one-by-one)
6. Add Novell Colors by giving them a name and entering RGB values
7. Click “OK”

# Graphics & Typeface



## Note:

**Icons/Lines:** This presentation refresh simplifies the current template and pushes focus on the content being presented. The icon library will continue to be utilized, but a refresh will be noticeable with the addition of the “Bubble” set of icons, and a subtle color shift. These icons are created to provide a professional, consistent look. When these icons are used sparingly, and in direct relation to the content on the slides, our presentations will communicate and work more effectively.

**Typeface:** Arial has been selected as the new typeface for all Novell communications. The following were considered.

1. Our typeface needs to be designed to carry information quickly to the reader.
2. It needs to be usable for Novell employees in company correspondence and presentations, as well as for outside vendors for marketing and promotion.
3. It needs to easily function on the Linux, Windows and Macintosh platforms.
4. And finally, Arial was created for these exact purposes.

Download Icon Library at: <http://innerweb.novell.com/brandguide>

## How to Add Novell Icons to OpenOffice Gallery:

1. Go to the “Tools” menu
2. Select “Gallery”
3. In the Gallery window select “New Theme...”
4. With the “General” tab active name your new theme (ie.Red flat)
5. Select the “Files” tab.
6. Select “Find Files...”
7. Find the downloaded folder containing the icons named and click “Select”
8. Select “Add All” and then “OK”
9. Repeat for all icon groups