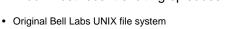


3

5

wska@cs.washington.edu Allen Center 570



- a simple yet practical design

- exemplifies engineering tradeoffs that are pervasive in system design
- elegant but slow
- · and performance gets worse as disks get larger

- solves the throughput problem

 - · aggressive caching
 - · awareness of disk performance details

11/15/2009

© 2009 Gribble, Lazowska, Levy, Zahorian

Both are real dogs when a crash occurs

- · Buffering is necessary for performance
- Suppose a crash occurs during a file creation:
 - 1. Allocate a free inode 2. Point directory entry at the new inode
- In general, after a crash the disk data structures may be in an inconsistent state
 - metadata updated but data not
 - data updated but metadata not
 - either or both partially updated
- fsck (i-check, d-check) are very slow
- must touch every block
 - worse as disks get larger!

© 2009 Gribble Lazowska Levy Zaborian

Journaling file systems

2

4

- Became popular ~2002
- There are several options that differ in their details - Ext3, ReiserFS, XFS, JFS, ntfs
- Basic idea

11/15/2009

- update metadata, or all data, transactionally
- "all or nothing"
- if a crash occurs, you may lose a bit of work, but the disk will be in a consistent state
 - · more precisely, you will be able to quickly get it to a consistent state by using the transaction log/journal – rather than scanning every disk block and checking sanity conditions

© 2009 Gribble Lazowska Levy Zaborian

Where is the Data?

- · In the file systems we have seen already, the data is in two places:
 - On disk

11/15/2009

11/15/2009

- In in-memory caches
- · The caches are crucial to performance, but also the source of the potential "corruption on crash" problem
- The basic idea of the solution:
 - Always leave "home copy" of data in a consistent state
 - Make updates persistent by writing them to a sequential (chronological) journal partition/file
 - At your leisure, push the updates (in order) to the home copies and reclaim the journal space

© 2009 Gribble, Lazowska, Levy, Zahorjan

