



200TB Linux based
Disk servers

Outline



- Can it be done?
- Problems to be solved.
- The future



Can it be done?

- It's the secret sauce!



Problems

- Several problems have to be solved. Some are simple, some are technical in nature.
- Linux client problems
- Linux server problems



Client Problems



- No real problems
 - Minor details
 - filesystems larger than 2TB create problems with 'stat' on disk space. (any filesystem larger than 2³¹-1*blocksizes)
 - Problem exists in all 32 bit Linux's, even in today's 2.4 kernel.
 - LFS (Large File Summit) support doesn't fix this problem. It's in the Linux VFS layer.
 - Major details
 - Only supports 128 NFS mounts at one time.

LAWRENCE BERKELEY NATIONAL LABORATORY



Server Problems



- Current 32bit Linux problems (v2.2 and v2.4)
 - No filesystem larger than 2³¹-1*blocksizes supported (~2TB)
 - Even problems with volumes with more than 2³¹-1*blocksizes bytes in size.
 - Changing block size doesn't help much
 - V2.4 supports NFS3, and LFS.
 - V2.4 supports 64 bit file handles.
 - V2.2 has v3 support, but no LFS support.

LAWRENCE BERKELEY NATIONAL LABORATORY



The Good news!



- Multiple 2TB volumes and filesystems are supported.
- Large client counts in server in v2.4 – no limits that I know of!
- Automounts work better in v2.4, including direct mounts.

LAWRENCE BERKELEY NATIONAL LABORATORY



200TB or bust?



- Well, we could try build 2TB servers.
 - Way too many servers – 100 servers? NUTS!
- Just what is a good number?
 - Let's say 10. That's 20TB per server, which is easy to do today (20,000GB/150GB = ~140 disks.)
 - At this number of disks, a complete FC based SAN behind each server for the volume sets.
 - Multiple GigE interfaces into a switch, either trunked or multihomed.

LAWRENCE BERKELEY NATIONAL LABORATORY



Can it be done?



- Yes, BUT
 - Small volume sizes (in relationship to total capacity)
 - Logical volume management **must** be used.
 - Journaling filesystems also.
- Might consider using ia64 systems for servers, that raises the size of the volumes, but clients still could have problems.