

MathWorks: A Case Study in NAS

Ira Cooper
The MathWorks Inc.

Who?

- ❑ MathWorks develops MATLAB™ and Simulink™.
 - ❑ Including ~80 Toolboxes and Blocksets together!
- ❑ We are a company of ~2500 people (~1000 developers) across many sites:
 - ❑ United States – HQ
 - ❑ France
 - ❑ Germany
 - ❑ Japan
 - ❑ India

- MATLAB™ and Simulink™ offer products supporting:
 - Model Based Simulation, Design & Verification
 - Finance
 - Statistics
 - Embedded Code Generation
 - Symbolic Mathematics
 - And many more areas

What?

- ❑ Extremely large codebase
- ❑ Builds can take 24hrs+ for a complete sterile build.
- ❑ Linear test time for all our tests is in the thousands of hours.
- ❑ MATLAB™ integrates and tests with many 3rd party products.
- ❑ MATLAB™ uses the file system as its namespace!

- ❑ Build and Test Lab! (BaT)
 - ❑ Helped debug SMB2 implementations.
 - ❑ Batch system, runs 24/7/365.

- ❑ Thousands of cores of compute power:
 - ❑ Mac OS X
 - ❑ Windows 32/64 bit
 - ❑ Linux 32/64 bit

Why Storage?

- ❑ Storage was holding up the company's progress.
 - ❑ Storage costs equaled compute node costs.
- ❑ Performance:
 - ❑ We need 100k+ ops for hours from a fileserver.
 - ❑ The more ops, the faster the jobs run, the more code goes into the product.
- ❑ Stability:
 - ❑ Over 24 hour build time

Why Storage 2?

- ❑ Bug Fixing and Verification:
 - ❑ Have to confirm the bugs are not product issues.
 - ❑ 90% of the fight is identifying the bug.
 - ❑ The other 90% is convincing the vendor to fix it.
- ❑ Now we can work these issues as needed.
 - ❑ If not, we can at least give very precise bug reports.

- ❑ High reliability:
 - ❑ To support the twenty four hour builds
- ❑ Performance:
 - ❑ 100k+ mixed ops per fileserver is not unheard of.
- ❑ Cost:
 - ❑ “NASCAR for the price of a Corolla.”

Design Factors 2

- ❑ Huge numbers of metadata operations.

- ❑ So we can bias our file servers to our needs:
 - ❑ Large amounts of memory to cache metadata
 - ❑ More servers, because they are cheaper!
 - ❑ Don't have to run them as hard
 - ❑ Close to the limits == Problems!

- ❑ High Availability:
 - ❑ There is none!
- ❑ “Deal with the Devil”:
 - ❑ Each server will go down one day a year, and we can’t say which.
 - ❑ The system will go twice as fast and cost much less.
 - ❑ For a batch system, this can be a great deal!

Design Choices 2

- ❑ No HA means simple servers.
 - ❑ For multi-protocol file servers

- ❑ It also allowed us to go from POC to production in about 6-8 months.

Design Choices 3

- Timely support is critical:
 - It can be provided in house.
 - We know and understand the priority of our own problems better than any vendor can.

What solution?

- ❑ Nexenta Core + Samba 3.6-GIT
- ❑ SuperMicro servers
- ❑ To say more...
 - ❑ 72gb+ RAM
 - ❑ 3 L2ARC SSDs
 - ❑ 2 ZIL SSDs
 - ❑ 19 10k drives
 - ❑ ... all in one 2U server.

Why Nexenta?

- ❑ OpenSolaris is a very solid OS.
 - ❑ Nexenta Core also acted more like Linux.
- ❑ ZFS is a very solid filesystem:
 - ❑ Raid – RaidZ, RaidZ2, Mirror
 - ❑ Snapshots
 - ❑ SSD Read Caching – L2ARC
 - ❑ Not quite “tiering”, but cost-wise that’s fine!
 - ❑ SSD Write Caching – ZIL
 - ❑ Required due to heavy NFS traffic

Why Nexenta 2?

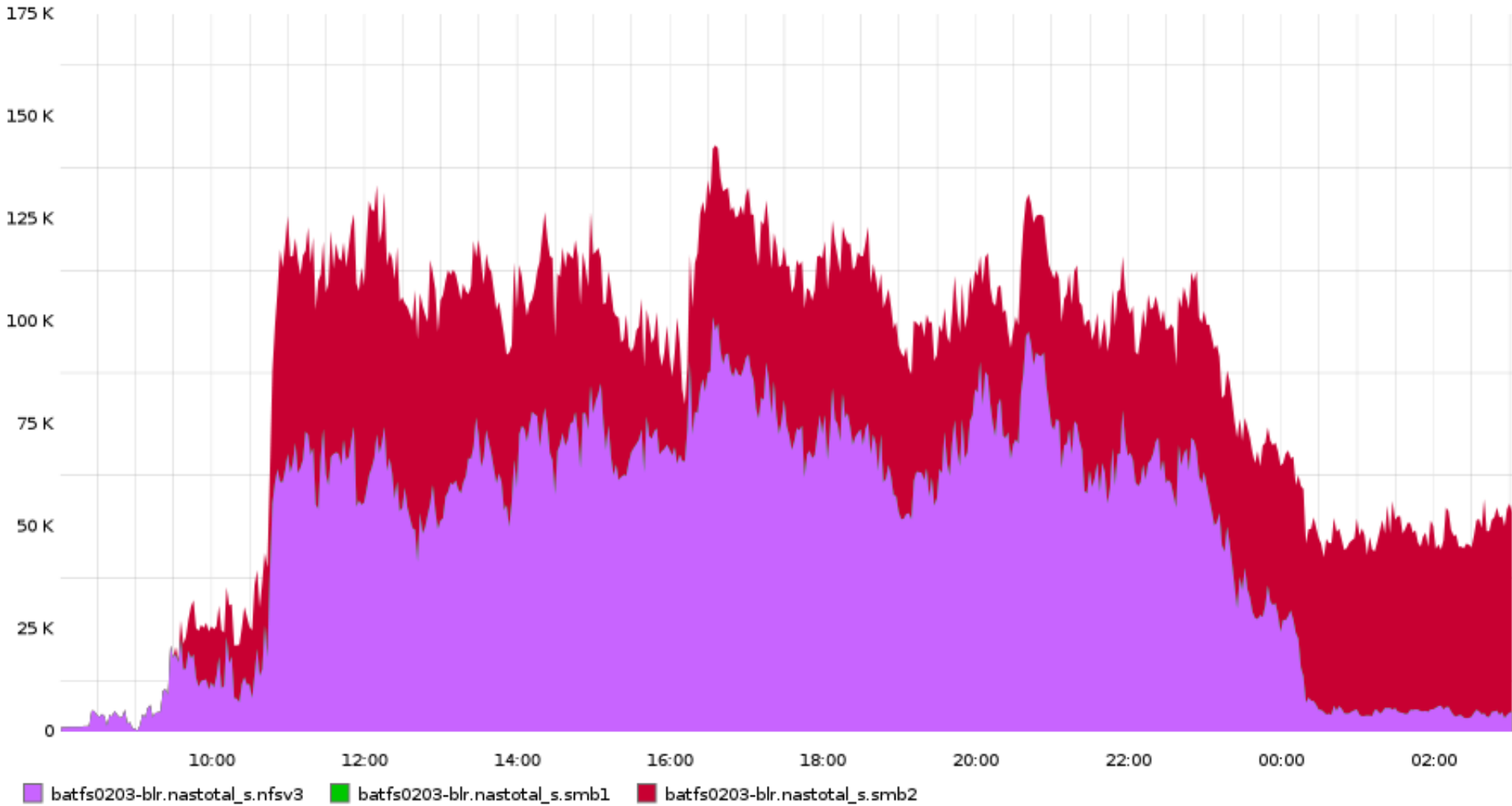
- Nexenta is doing what we are doing!
 - ZFS
 - NFS
 - On SMB we disagree:
 - Kernel CIFS vs. Samba
- Always better to work with people whose interests align with yours!

Why Samba?

- ❑ Kernel CIFS had issues in our AD environment.
 - ❑ Also, it did not have SMB2.
- ❑ Likewise is missing a key feature: Notify.
- ❑ Samba has a very mature codebase.
- ❑ Samba also has a very strong community.
 - ❑ This leads to more features and bug fixes.

Results!

batfs0203-*blr OPS



Results 2!

- ❑ Management is more understanding of the effort involved because we are more hands-on.

- ❑ They want the problem to be in our servers.
 - ❑ They are confident we can fix it!

Example Problem: C/N

- ❑ SMB2.002.
 - ❑ Compounded Create + Notify.
 - ❑ Every so often, servers wouldn't reply correctly.
 - ❑ Causing disconnects
 - ❑ Causing our builds to fall over
 - ❑ We were able to diagnose the issue.
 - ❑ However, vendors wouldn't take notice.

Example Problem: C/N 2

- What can we do?
 - Even when we got a solution from a vendor, it might just fix the bug “sometimes”.
- We want to roll out SMB2 badly to speed up builds.
 - A large speed up in our builds is critical to our business.

Example Problem: C/N 3

- ❑ Eventually, we decided to do it ourselves.
 - ❑ We worked in house to develop an awful prototype patch.
 - ❑ The Samba team rewrote that patch.
 - ❑ We worked with the Samba team to QA the final version that went into Samba.

Example Problem: C/N 4

- ❑ Samba's fix led to the issue being recognized throughout the industry.
- ❑ Internally, we realized “Hey, just like in MATLAB, finding and isolating the bug is often over 90% of the fight!”
 - ❑ So why not just do it!

Example Problem: Groups

- ❑ Solaris only allowed a user to be in only 32 groups.
 - ❑ If `setgroups` was called with more, the process was killed!
 - ❑ Samba thought this was a security issue: people should be in the groups AD says.
 - ❑ The security issue was not a concern for us.
 - ❑ We patched our own version and went about life.

Example Problem: Groups 2

- ❑ A key thing about open source:
 - ❑ Nobody can stop you from doing what you need to do to make it work!
 - ❑ Work together as you can.
 - ❑ Agree to disagree as you must.

Example Problem: ZFS Panic

- ❑ After one of the updates to Nexenta Core we started getting a kernel panic.
 - ❑ Clearly we couldn't ship with it in place.
 - ❑ We were able to take crash dumps.
 - ❑ Used the source + Mercurial to find out what happened.
 - ❑ Patched it locally.
 - ❑ We felt confident, but ZFS is hairy.

Example Problem: ZFS Panic

- ❑ We sent Nexenta a patch, and asked them to verify it.
 - ❑ It went into their shipping products!
- ❑ They also asked us to QA a better fix for this issue.
 - ❑ We did and now use their fix.
- ❑ Work with people; make it worth their time.

- ❑ No HA does mean that from time to time things do go wrong, and we have more to clean up.
- ❑ No direct vendor support:
 - ❑ You broke it, you own both halves.
- ❑ Learning how to build a NAS device in 6 months was a challenge.

- ❑ The customer is always right:
 - ❑ Especially when he is right down the hall from me and paying my check directly!
- ❑ Working with the open source community has been a wonderful thing.
- ❑ The servers have been a great success.
 - ❑ They have enabled the growth of the BaT lab and the company.

Questions?



Thank You for Attending!

