OpenSolaris™ NFS Performance

Alain van Hoof

February 3, 2010
Research Question

\begin{quotation}
How can the performance bottlenecks be monitored and identified on an OpenSolaris OS NFS server.
What are realistic load simulations and create a base-line.
\end{quotation}
NFS layer and ......
Inspecting the layers

Client Linux

NFS Client

TCP/IP

TCP/IP

NFS Server

VFS Layer

ZFS FileSystem

ZIL

L2ARC

Physical Devices

Server OpenSolaris

Dtrace Probes

mib:ip:ip_
nfs3:nfsrv:nfs3_
tszinfo:genunix:tcp_
fbt:zfs:arc_
fbt:zfs:zil_lwb_write
io:genunix:start

nfsstat -c
etstat
tcpdump/wireshark
netstat
nfsstat -s
fsstat
zilstat
arcstat
zpool iostat
iostat

4
Benchmarking

Macro benchmarks

*real-world workload*

Trace benchmarks

*replay recorded real-world scenario*

Micro benchmarks

*one or two operations*
Filebench

Macro benchmark

Supported by SUN™

fileserver: SPECsfs workload

IOzone

Micro benchmark

A benchmark UvA-IC is familiar with

Random Read/Write
Auto-pilot

Automate the running of benchmarks
Avoid user errors (typo’s)
Automatic mount and unmount filesystems
Calculate confidence level and run more test
to be more confident
Test Setup

VMware 01
- LoadClient 01
  - autopilot
  - filebench

VMware 02
- LoadClient 03
  - autopilot
  - filebench
- LoadClient 04
  - autopilot
  - filebench

LoadClient 02
- autopilot
- filebench

TestClient
- Test Results
  - auto-pilot
- filebench
- IOzone

NFS Server

Server Metrics Log
Benchmark & load generation

Run benchmark on TestClient while load is generated on LoadClients

Real-world benchmark is also real-world load generator
Meanwhile on the NFS server

12 Dtrace probes are logged during testing
Output: Every 10 seconds
Probes triggered/second
Preliminary Test

A simple test to learn about behavior of the Dtrace probe

“Accidentally” Identified NFS bottle-neck

Changed number of NFS daemons

Positive result
Dtrace write probes

Before

After

Time

Awr/32
Bwr
Cwr
Dwr
E
G/10
Results of Test

Filebench and IOzone show performance loss when load increases.

Filebench more close to “user experience”

IOzone tests are not within the confidence interval (95%) after 30 runs.
filebench

MB/s

Load
IOzone Random Write

MB/s vs Load

Load:
- 0
- 1
- 2
- 3
- 4
- 8
- 16

MB/s:
- 0
- 5
- 10
- 15
- 20
- 25
- 30
- 35
- 40

Data points indicate the performance of IOzone Random Write at different load levels.
ZIL write probe
ZFS and ZIL

All NFS writes are Synchronous and pass the ZIL
Improve performance

Place ZIL on fast device
Solid State Disk (SSD)
Well known practice
Results - ZIL on SSD

Filebench and IOzone show performance increase

When the load increases performance decreases

IOzone tests are not within the confidence interval (95%) after 30 runs
Dtrace write Probes

ZIL write probe
Conclusion

Using Dtrace probes, performance bottlenecks on the server can be identified, and a baseline can be created.

A realistic load can be generated using the filebench - fileservserver personality

Filebench test created a confident baseline (unlike IOzone)
Questions?

Thanks to:

Jeroen Roodhart

Auke Folkerts