Performance Analysis
Benchmark of the sgi TP9400

William Julien
Engineering Operating Systems
Boeing Shared Services
P.O. Box 3707 MC 7J-04
Seattle, WA 98124-2207 USA
julien@sdc.cs.boeing.com
Our humble Origin

✓ 8 CPU Origin 2000
  ✓ Primarily cpu intensive codes.

✓ I/O performance not a priority

✓ Clarian raid storage device
  ✓ not configured with write cache
  ✓ initially configured striped xlv across two sub-volume luns

✓ Eliminated the sub-volume stripe and reconfigured as an 8+1 raid5
  ✓ performance improved 100%. 
Origin 2000 Clarian Raid
Diskperf Direct I/O

- subvolume write
- subvolume read
- 8+1 write
- 8+1 read

Request Size

MB/Sec
Test and evaluation of the Origin 3800 and tp9400

✓ Origin 3800
✓ 60p prod
✓ 4p test
✓ cxfs
✓ potential I/O intensive codes
✓ Extensive system QA performance test
✓ tp9400 test

✓ Methodology
✓ diskperf -D -W
✓ local application

```c
fd = open(streamName, "w", O_WRONLY | O_CREAT);
for ( block = 0; block < Nblocks; block++) {
    n = write(fd, buffer, blockSize);
}
```

Test platform
✓ Origin deskside
✓ 4 cpu Onyx
✓ brocade switch
✓ tp9400
✓ 2 front end hubs
✓ 4 back end hubs
✓ 2 4+1 w/2 lun stripe
TP9400 Filesystem Test
4k blk - 1M Stripe unit - 256k segment size
Observations and Actions Taken

✓ Observations
  ✓ Tests on CXFS had similar performance.
  ✓ I/O rate was linear to the number of streams running.
  ✓ Our applications are primarily write, so read ahead cache was not enabled.

✓ Actions
  ✓ Enabled Command Tag Queuing.
  ✓ scsifo -d
  ✓ The primary paths on the LUNs defaulted to the same controller!
  ✓ Set primary paths with failover.conf
  ✓ Rebuilt the filesystem with a 65536 block size and a 2048 stripe size.
TP9400 Performance after Tuning

Via the Brocade Switch

- Fwd_write
- Fwd_read
- Bwd_write
- Bwd_read

Running at the limit of the two tp9400 front-end hubs
There be dragons here
System became unresponsive by running 10 large diskperf (or dd) streams

- Case# 2174429
- floods kernel buffer cache
- maci sort routine
- new processes can be created, but appear to hang on exit
- I/O throughput drops to less than 2 MB/sec
Conclusions

✓ Test your I/O performance.
   ✓ you may be surprised what you find!
   ✓ diskperf -D -W -c3000 testfile

✓ Enable command tag queuing.
✓ Check your data paths.
   ✓ scsifo -d | grep P
   ✓ configure failover.conf

✓ Use as many spindles as possible.
   ✓ 18gig vs 36gig drives.
   ✓ stripe over as many luns as you can.
   ✓ do not subpartiton your luns.

✓ 100MB/sec per host minihub (gbic)
   ✓ 4 minihubs for maximum performance
Questions?

William Julien
Engineering Operating Systems
Boeing Shared Services
P.O. Box 3707 MC 7J-04
Seattle, WA 98124-2207 USA
julien@sdc.cs.boeing.com