The world is a file

nfs://industry.conf

NFS INDUSTRY CONFERENCE

Analyzing NFS Client Performance with IOzone

Don Capps
Performance Architect
HP
capps@iozone.org

Tom McNeal
Independent Consultant
TMCN Consulting
trmcneal@attbi.com

The world is a file
nfs://industry.conf

NFS INDUSTRY CONFERENCE

Benchmark Overview

Characteristics of IOzone Activities

Load Generation

- File System I/O requests
  - File sizes vary from 64K to 512M
  - Record sizes vary from 4K to 16M
  - Each increase doubles previous size
  - Large file system calls supported
- System variants supported
  - Memory mapped files
  - fread(), fwrite()
  - pread(), pwrite()

Sequential Reads/Writes

- Reads & Rereads
- Writes & Rewrites
- Backwards sequential read
- “Stride” read
  - Uses constant intervals for sequential reads from beginning to end

Other Reads/Writes

- Randomized Reads/Writes
- Record Rewrite (from offset 0)
- fread() – Reads and Rereads
  - Serialized, Buffered & Blocked IO
- fwrite() – Writes and Rewrites
  - Serialized, Buffered & Blocked IO

The world is a file
nfs://industry.conf

NFS INDUSTRY CONFERENCE

Recommended Variants for NFS Clients

```
./iozone  –azc  –U  /mnt/testdir  –f  /mnt/testdir/testfile
```

- All tests, all record sizes
- Commit time included in measurements
- IO targeted at mounted file
  - Unmount clears out caches between tests
  - Target file specified in mounted directory

The world is a file

nfs://industry.conf

NFS INDUSTRY CONFERENCE

Benchmark Results

Characteristics of IOzone Reports

Graphical Reports

```
./iozone  –R  -b  exceloutput.xls  > logfile
```

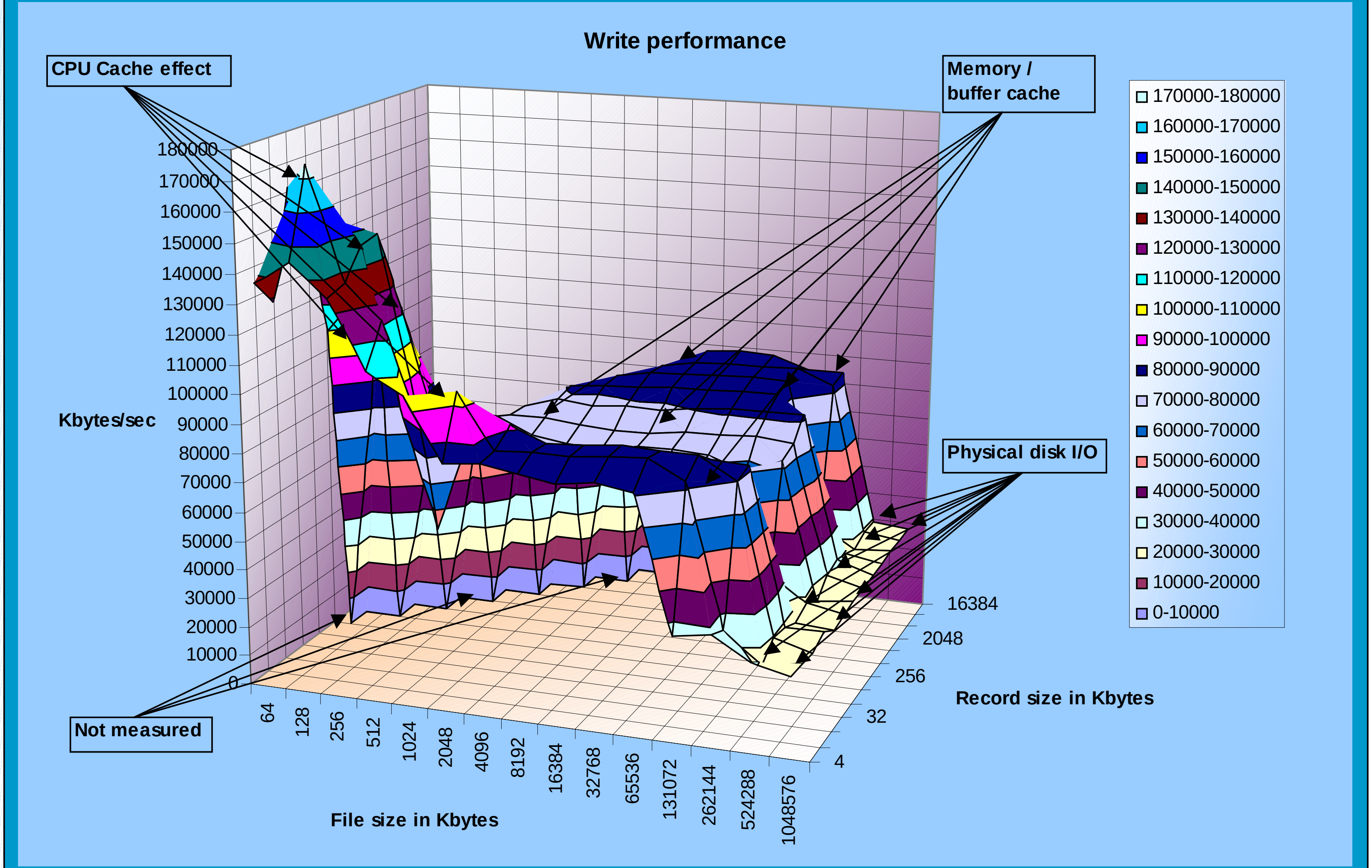
- Generate Excel output text
- Named file has graphs and data
  - 3D Surface Charts for all tests
  - Includes text output used for graphs
- Standard output sent to log file
  - Generally useful for debugging problems

The world is a file

nfs://industry.conf

NFS INDUSTRY CONFERENCE

Surface Plot Graphs


Write performance

CPU Cache effect

Memory / buffer cache

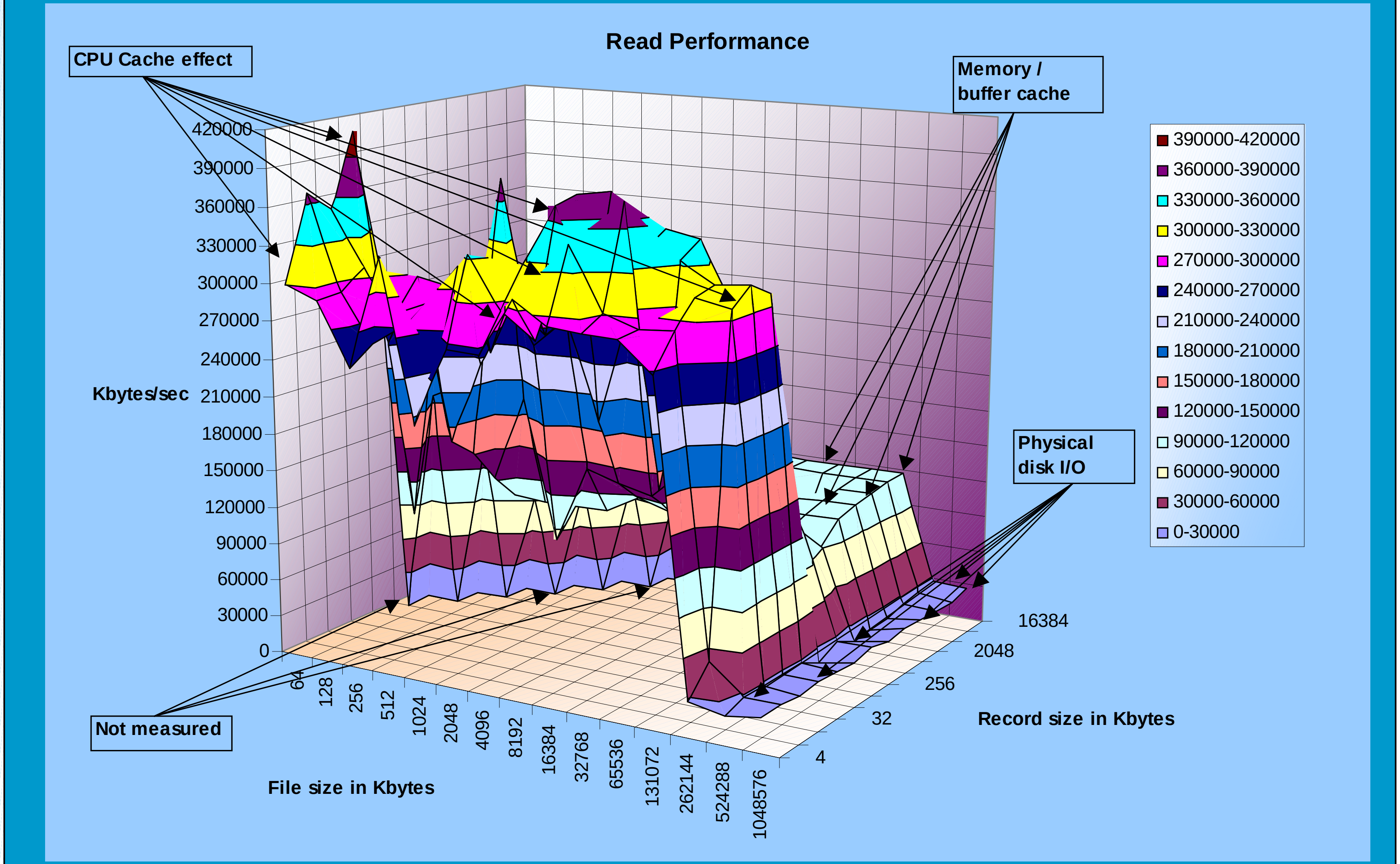
Physical disk I/O

Not measured

nfs://industry.conf

NFS INDUSTRY CONFERENCE

Surface Plot Graphs II


Read Performance

CPU Cache effect

Memory / buffer cache

Physical disk I/O

Not measured

System & Network Variants

Environmental Control and Interactions

System Level Variations

-p, -P #, -l n, -S size, -L size

- SMP Issues
  - Processor cache purges
  - Processor affinity (for a given # of cpus)
  - Lower bound of number of cpus
- Cache Management
  - CPU Cache size
  - CPU Cache line size

System parameters

- Client BIOD Daemons
- Server NFSD Daemons
- Number of file system nodes
  - rnode/inode/vnode/file handles
- Directory Name Lookup Cache
- Network buffer sizes

File System Variations

-o, -W, -e, -g #

- O_SYNC file option for all tests
- File locking required for all IO
- Flush timings included
  - fsync() and fflush()
- Large file offsets
  - File system calls determined at make time
  - Alternate max file size may be specified

File System Variations II

-B, -D, -G, -H n, -k n

- Memory mapped file IO
  - mmap() interface
  - MS_ASYNC or MS_SYNC usage available
- Posix asynchronous IO

Network Variations

- UDP/TCP Protocol
- Client transfer sizes
- Network speed, duplex settings
  - Autonegotiation is often “interesting”
- IP issues
  - Jumbo frames with gigabit ethernet
  - Stream heads, Socket buffer sizes

Clustered Clients

Managing and Measuring a Cluster

Client Specification

-+m filename

- Clients specified in a file
- Clients must be accessible
  - Remote shells enabled through .rhosts
- DNS ®
- IOzone revision 3.128 or later
- Stonewalling helpful (removed by –x)

“Stonewalling”

- Client tests initiated in tandem
  - All clients kept equally busy
- When one finishes, they all finish
  - Tests halted when the first client completes
- Emulates high performance parallel processing clusters
  - Beowulf clusters at LLNL, PNNL, Los Alamos

Summary

Examples and References

Summary

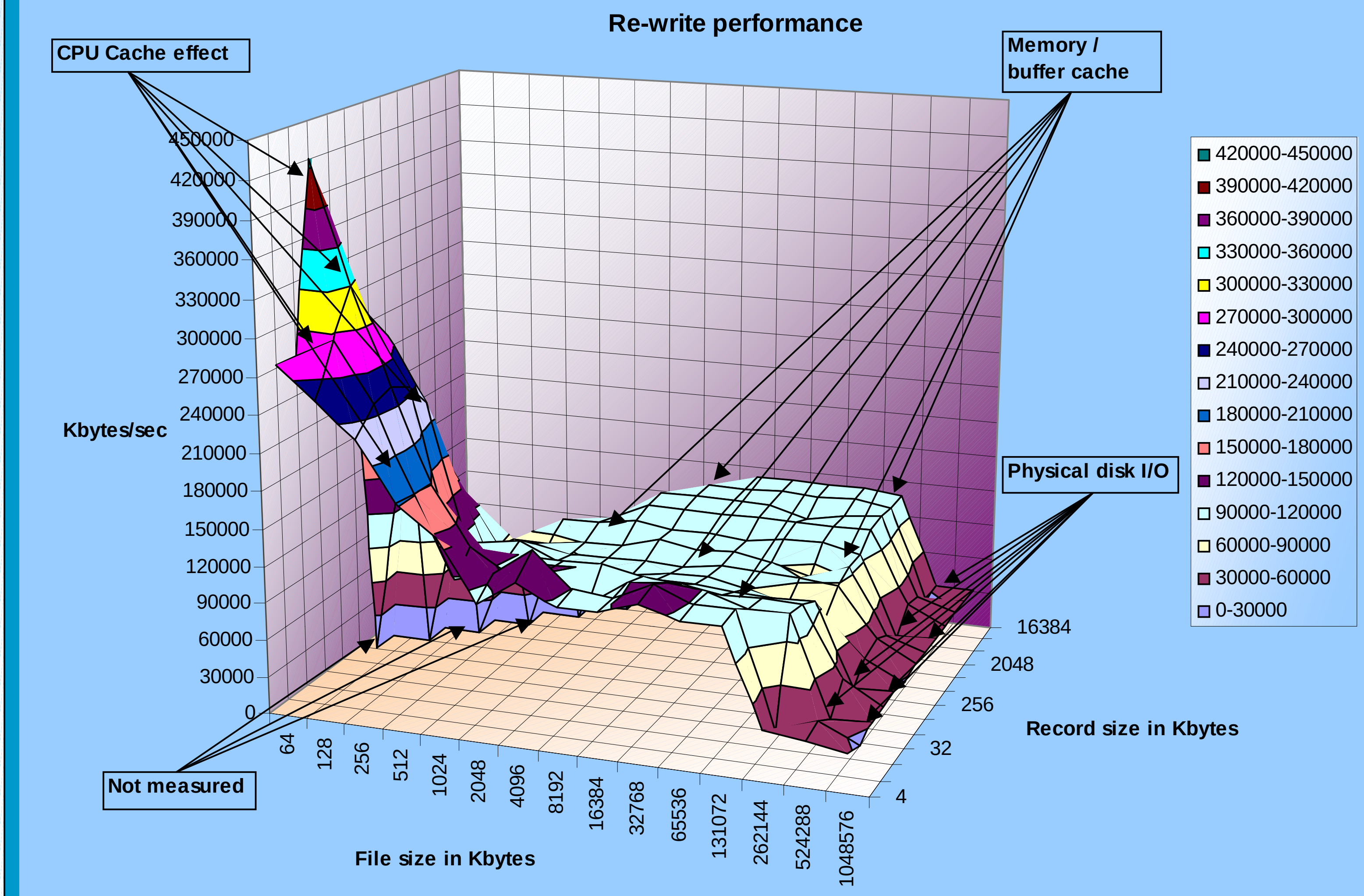
- NFS Client measurement standard

```
./iozone –azcR –U /mnt/testdir –f /mnt/testdir/testfile \
 -b exceloutput.xls > logfile
```

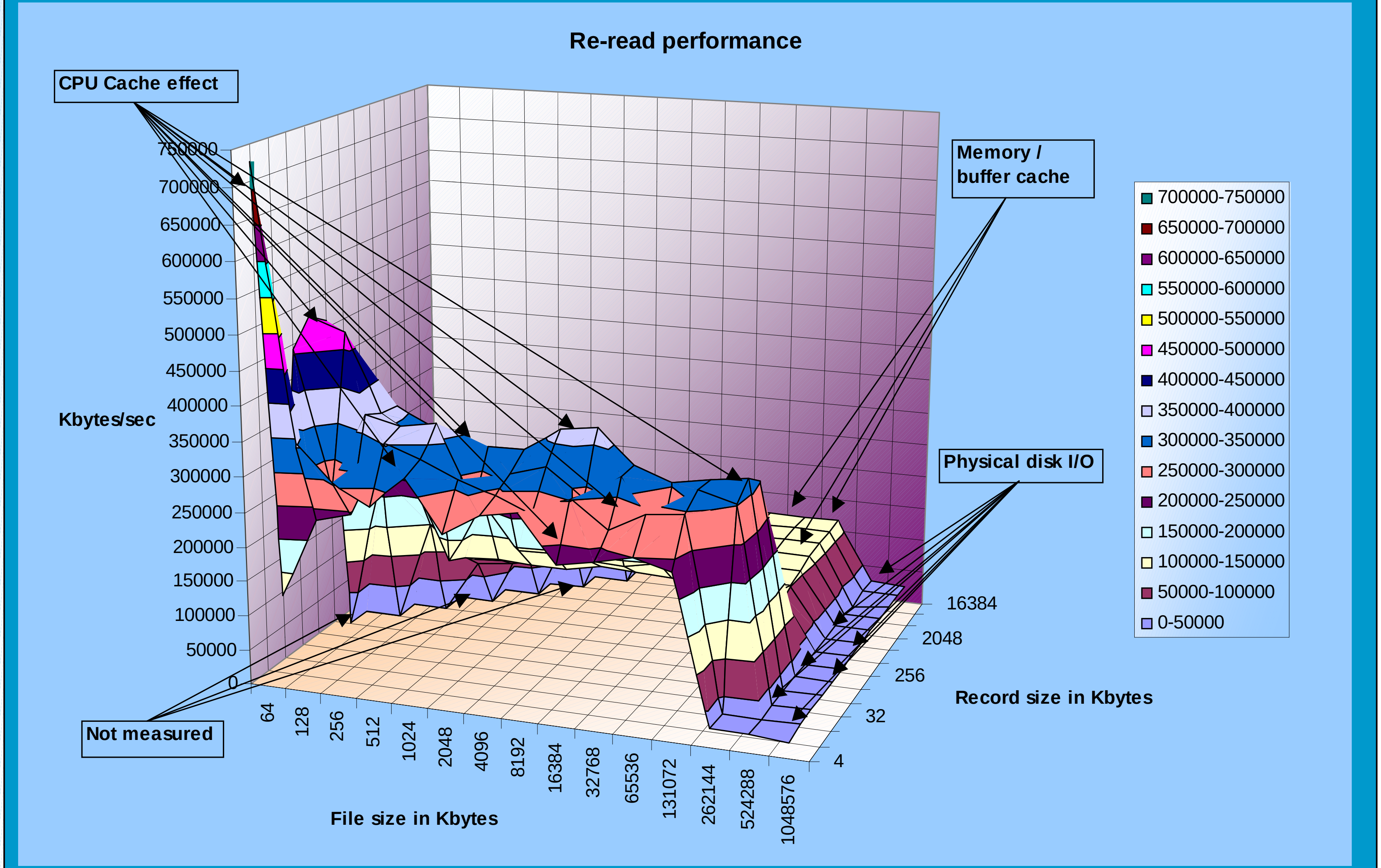
- Gather standard data first
  - What is right for your client?
- Review Variations and Features
  - Review System, FS, and Network setup
  - Start tuning, playing, tuning, playing….

NFS INDUSTRY CONFERENCE

Rewrite Graph



Reread Graph



Re-read performance

CPU Cache effect

Memory / buffer cache

Physical disk I/O

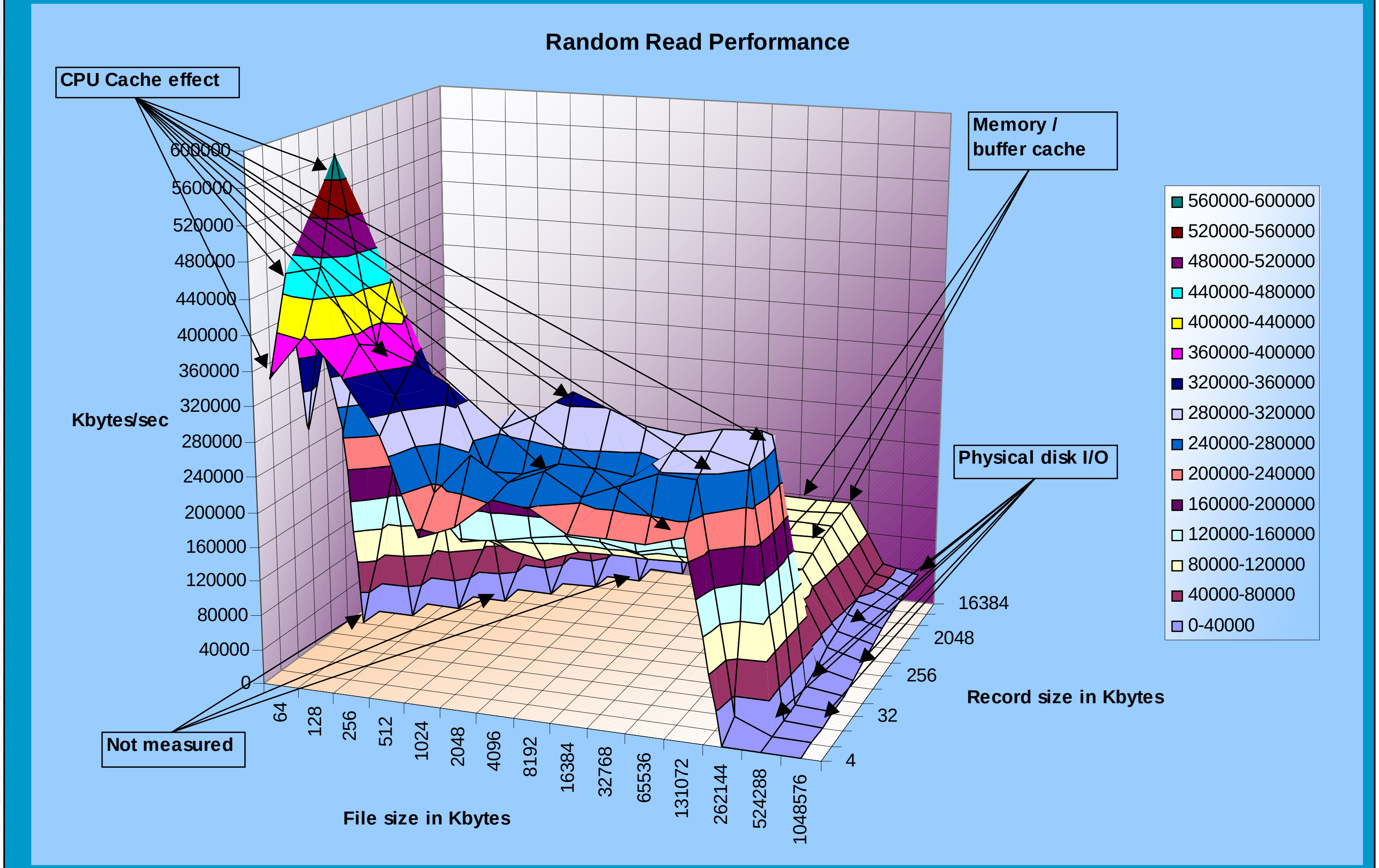
Not measured

The world is a file

nfs://industry.conf

NFS INDUSTRY CONFERENCE

Random Read Graph



Random Read Performance

CPU Cache effect

Memory / buffer cache

Physical disk I/O

Not measured

References

- http://www.iozone.org
- http://www.iozone.org/src/current
  - Contains 8K vs. 32K Transfer Size graphs
- http://www.sourceforge.net/projects/nfstestmatrix
  - Includes functional tests, destructive tests, and benchmarks for Linux systems
- http://www.mclx.com/orph
  - Linux Server performance review (late 2001)