

Exercise Introduction

This exercise sheet accompanies the IO500 Benchmark HPCSA. You will benchmark the NFS-mounted home directory from both the headnode and compute nodes to compare storage performance across different access patterns using MPI.

Total Time: 35 minutes (4 tasks + setup)

Activity	Time	Notes
Setup	5 min	SSH, modules, NFS verification
Task 1: Installation	10 min	Build IO500 on headnode
Task 2: Headnode Benchmark	10 min	Benchmark NFS from headnode
Task 3: Compute Node Benchmark	5 min	Run via SLURM job on 2 nodes
Task 4: Comparison	5 min	Analyze performance difference
Total	35 min	

Prerequisites:

- User account from earlier exercise.
- NFS home directory at `/nfs/home/$USER`.
- Access to both headnode and compute nodes.

Learning Goals:

- Measure NFS storage performance using IO500.
- Compare performance from headnode vs compute nodes.
- Understand network storage bottlenecks.
- Analyze scaling and I/O balance on shared filesystems.

Contents

1 Setup and Verification (5 minutes)	2
Task 2: Task 1: IO500 Installation and Build (10 min)	2
Task 3: Task 2: Headnode NFS Benchmark (10 min)	3
Task 4: Task 3: SLURM Job Benchmark from Compute Nodes (5 min)	5
Task 5: Task 4: Comparison and Discussion (5 min)	5
Optional Task 6: Advanced Exploration (Optional) (10 min)	6

1 Setup and Verification (5 minutes)

Verify your environment and confirm that your NFS home directory is accessible.

Connection and Environment Check

1. Connect to the headnode:

```
ssh <your-username>@<headnode-address>
```

2. Verify NFS mount:

```
echo $HOME
# Should show: /nfs/home/<your-username>
ls -la $HOME
df -h $HOME
mount | grep nfs
```

3. Load the MPI module:

```
module avail
module load openmpi    # or module load mpi
```

4. Verify MPI compiler and SLURM setup:

```
mpicc --version
sinfo      # Check available compute nodes
```

5. Prepare working directory:

```
cd $HOME
mkdir -p io500-hpcsa
cd io500-hpcsa
pwd      # Verify /nfs/home/<user>/io500-hpcsa
```

Hints

- Confirm your NFS mount is active: `mount | grep nfs`.
- Check available space: `df -h /nfs/home`.

Task 2: Task 1: IO500 Installation and Build (10 min)

Objectives

- Clone and build IO500 on headnode.
- Verify MPI linkage and binary accessibility from compute nodes.

Steps

1. Change to your working directory:

```
cd $HOME/io500-hpcsa
```

2. Clone the official repository:

```
git clone https://github.com/IO500/io500.git
cd io500
git checkout io500-sc23
```

3. Set compiler environment:

```
export CC=mpicc
export CXX=mpicxx
```

4. Build IO500 (approx. 3–5 minutes):

```
./prepare.sh
```

5. Verify build:

```
ls -lh io500 bin/
./io500 --help | head
```

Verification

- Test executables:

```
test -x ./io500 && echo "IO500: OK"
test -x ./bin/ior && echo "IOR: OK"
test -x ./bin/md-workbench && echo "MDTest: OK"
test -x ./bin/pfind && echo "pfind: OK"
```

Hints

- If mpicc not found, load MPI: `module load openmpi`.
- Built binaries are on NFS, thus visible to all nodes.

Task 3: Task 2: Headnode NFS Benchmark (10 min)

Objectives

- Run IO500 on NFS from headnode.
- Establish baseline performance metrics.

Steps

1. Copy minimal config and edit paths:

```
cp config-minimal.ini config-headnode.ini
nano config-headnode.ini
```

Example minimal configuration:

```
[global]
datadir = ./datafiles
resultdir = ./results
```

```

api = POSIX

[debug]
stonewall-time = 60

[ior-easy]
API = POSIX
transferSize = 64k
blockSize = 256m
filePerProc = TRUE
run = TRUE

[mdtest-easy]
API = POSIX
n = 50
run = TRUE

[ior-hard]
API = POSIX
transferSize = 64k
blockSize = 128m
run = TRUE

[mdtest-hard]
API = POSIX
n = 50
run = TRUE

[find]
run = TRUE

```

2. Prepare directories:

```
mkdir -p datafiles results
```

3. Run test using two MPI processes:

```
mpirun -np 2 ./io500 config-headnode.ini
```

4. Find and record results:

```
ls -lt results/
```

Observation Table

Metric / Observation	Value / Notes
Runtime (sec)	
Bandwidth (GiB/s)	
IOPS (kIOPS)	
Errors / Warnings	
NFS responsiveness during run	

Table 1: Headnode Benchmark Results

Task 4: Task 3: SLURM Job Benchmark from Compute Nodes (5 min)

Objectives

- Run IO500 through SLURM using 2 nodes.
- Compare to headnode performance.

Steps

1. Create job script:

```
cat > io500_compute.sh << 'EOF'
#!/bin/bash
#SBATCH --job-name=io500-nfs
#SBATCH --nodes=2
#SBATCH --ntasks-per-node=2
#SBATCH --time=00:15:00
#SBATCH --partition=standard # change as per the partition name.
#SBATCH --output=io500_compute_%j.out
#SBATCH --error=io500_compute_%j.err

module load openmpi
cd $HOME/io500-hpcsa/io500

mpirun ./io500 config-headnode.ini
EOF
```

2. Submit job:

```
sbatch io500_compute.sh
```

3. Check job and outputs:

```
squeue -u $USER
cat io500_compute_*.out
```

Hints

- Compute nodes access IO500 binaries via NFS.
- Expect slightly slower results due to NFS network overhead.

Task 5: Task 4: Comparison and Discussion (5 min)

Compare Key Metrics

```
grep "Bandwidth\\|IOPS\\|TOTAL" results/*/result_summary.txt
```

Fill in Table 2:

Table 2: Headnode vs Compute Node Comparison

Metric	Headnode (2p)	Compute (4p)	Ratio
Bandwidth (GiB/s)			
IOPS (kIOPS)			
Total Score			
Runtime (sec)			

Discussion Questions

1. Does the compute-node run scale linearly ($\approx 2\times$ performance)?
2. Which component (IOR vs MDTest) scales better?
3. What factors limit NFS scalability?
4. What could be improved (network, cache, or NFS tuning)?

Optional Task 6: **Advanced Exploration (Optional) (10 min)**

This is a difficult **additional** task which will support your understanding in the topic.

Option 1: Analyze CSV Output

```
grep "ior-easy-write\\|ior-hard-write" results/*/result_summary.txt
head -20 results/*/ior-easy-write.csv
```

Option 2: Modify SLURM Configuration

Try:

```
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
```

Compare with two-node job performance.

Option 3: Network Observation

```
watch -n 1 'nfsstat -c'
ifstat -i eth0 1
mount | grep nfs
```

Portfolio (directory: 2/io500-hpcsa-results)

2/io500-hpcsa-results/headnode-results.txt	Copy of result_summary.txt (Task 2)
2/io500-hpcsa-results/compute-results.txt	Copy of result_summary.txt (Task 3)
2/io500-hpcsa-results/comparison-analysis.txt	Filled comparison table and discussion
2/io500-hpcsa-results/io500_compute_*.out	SLURM output log
2/io500-hpcsa-results/extra-analysis.pdf	(Optional) Extended study or plots

Further Reading

- IO500 Official Website: <https://io500.org>
- IO500 Documentation: <https://io500.org/documentation>
- Kunkel et al., “The Virtual Institute for I/O and the IO-500”, 2020.
- NFS Performance Tuning Guide: <https://nfs.sourceforge.net/>
- Course Resources: <https://hps.vi4io.org/teaching/ressources/start>

Appendix: Quick Reference Commands

Environment Checks

```
echo $HOME
mount | grep nfs
module list
sinfo
```

Monitoring

```
tail -f results/*/result.txt
nfsstat -c
ifstat -i eth0 1
iostat -x 2
```

Troubleshooting

```
cat io500_compute_*.err
scancel <jobid>
test -x ./io500 && echo "OK"
ls -la /nfs/home/$USER
```