NOTE 232 – Parallelization Test Results

Wes Young & Doug Roberts

May 8 2000

Contents

1 Introduction

We present (and document) preliminary test results using the AIPS++task pimager (a parallelized version of imager) on both a cluster and a DSM system. We did a computational test using the Clark CLEAN Algorithm and a parallel IO test. We ran the tests on the Origin 2000 at the NCSA in Urbana IL, the Origin 200 at the NRAO in Socorro NM and on the Linux cluster (roadrunner) at the Albuquerque High Performance Computing Center (AHPCC) at the University of New Mexico in Albuquerque NM.

2 Tests

The two tests are described as follows:

- Parallelized deconvolution over spectral channels using the AIPS++function bigimagertest to do a Clark CLEAN on a subset of the M33 data set supplied by Dave Westpfahl. specifically

```
niters := 50000; nthread := 5; processors := nthread-1;
print spaste('2kx2kx4 - ', niters, ' iterations - ', processors,
```

```
                    ' worker threads');
     bigimagertest(size=2048, cleanniter=niters, imagertask='pimager', nchan=4,
                    numthreads=nthread, forcenew=T);
```

with nthread being varied 2,3, and 5.

- A simple parallel iteration through the complete M33 uv-data set using the AIPS++function pimager.tryparread.

```
pim := pimager ("M33-A.MS", numprocs=i)
pim.tryparread("M33-A.MS", numloops=128)
```

Where numprocs was varied from 2,3,5, 9, 17, 33, 65, and 129. The run with numprocs set to 129 was only done on the Origin.

Both sets tests were run with and without collecting Pablo statistics. Pablo is the parallel I/O profiling library developed at UIUC. For more details about Pablo visit their web site at www-pablo.cs.uiuc.edu.

3   Results

3.1   Clark CLEAN Timings

The following table contains the timing data for the parallelized Clark CLEAN.

Processors	Roadrunner Time(s)	SGI Origin 200 Time(s)
1	1068	602,600,620
2	729,921	316,326,317
4	1645,1243,1157	208,214,211

Figure 1 plots the number of processors versus the median time to completion for the Clark CLEAN algorithm with 50000 iterations. Figure 2 shows the relative speedup of time versus processors for the median time to completion for the Clark CLEAN algorithm with 50000 iterations.

3.2   Parallel IO Tests

This table contains the timing data from the parallelized IO test. Note: timings on the Linux cluster with and without Pablo enabled were essentially the same.

| Processors	Roadrunner Time(s)	Normalized SGI Origin 2000 Time(s)	
		With Pablo Enabled	Without Pablo Enabled
1	3772,3815,3772	6932,5704,6127	5610,5613,5745
2	1922,1931,1914	2998,2914,2854	3080,2980,2830
4	1006,1013,1018	1547,1560,1500	1508,1505,1523
8	522,549,534	834,878,838	760,770,734
16	318,320,316	456,456,481	392,383,378
32	275,202,257	244,251,248	317,208,206
64	308,93,278	277,276,279	159,159,161
128			283,283,300
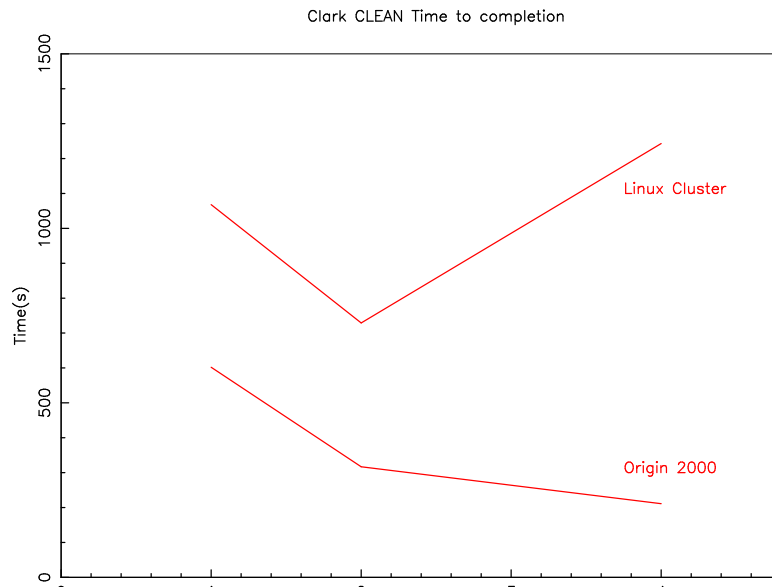
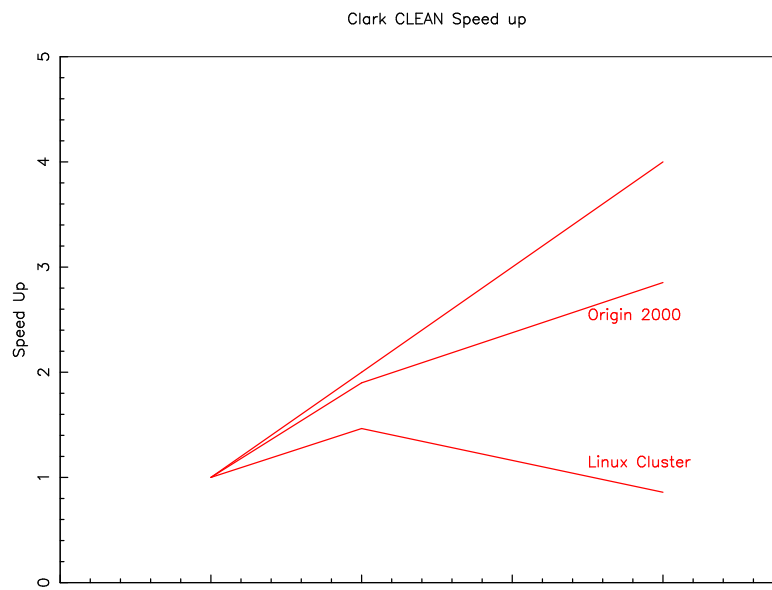Figure 1: Clark CLEAN time to completion
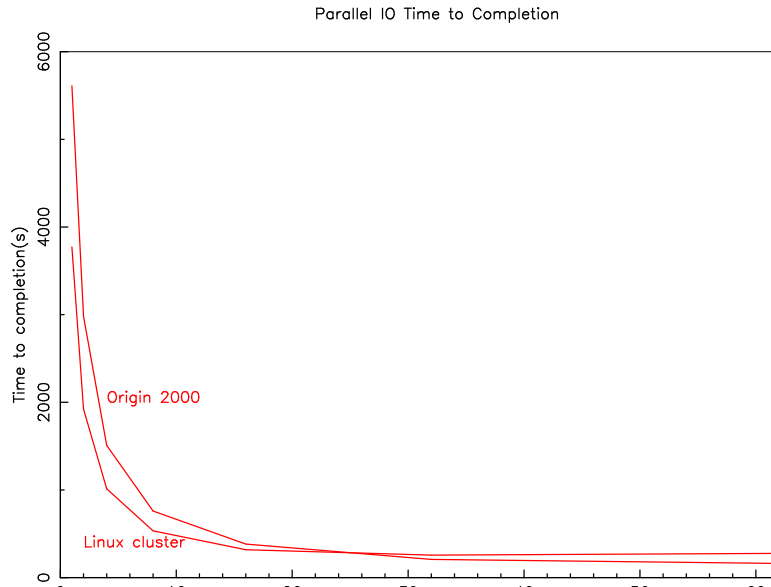

Figure 2: Clark CLEAN speedup

3

Figure 3: Parallel IO test time to completion

Figure 3 plots the number of processors versus the median time to completion for the parallel IO test. Figure 4 shows the relative speedup of time versus processors for the median time to completion for the parallel IO test.

Note: because of the heterogenous types of processors available at the NCSA the timings are adjusted to a normalized 175MHz R10000 chip. Absolute timings on the 225MHz chips were about 30% faster.

4   Comments

4.1   Clark CLEAN Results

The Clark CLEAN tests show a modest improvement as we increase the number of processors. Previous Clark CLEAN deconvolution tests on the Origin 2000 have shown good speed-up using 32 processors. The benchmark reported here needs to be repeated using more processors.

The Linux cluster shows no improvement as we increase the number of processors! This is a perplexing result and needs some addition investigation.

We encounter AIPS++table locking problems on the Linux cluster at times; further investigation is needed to identify the problem. This was not a problem on the Origin.
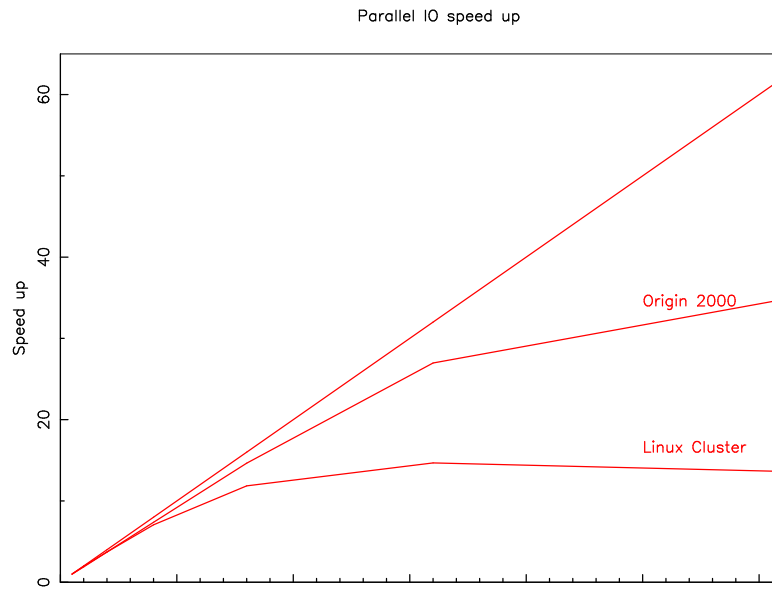
Figure 4: Parallel IO test speedup

4.2   IO Results

The roadrunner cluster is faster doing the parallel IO tests than the Origin with less than 32 processors.

The Origin show no improvement in performance past 32 processors on the IO tests when collecting Pablo statistics.