

The ClusterWorld Benchmark Project

There is a kind of joke in the cluster business. You can answer every question asked about clusters with one simple line: *It all depends on the application*. Remember this phrase and you have just become an instant expert on HPC clusters. Joking aside, there is quite a bit of truth in this answer. With cluster computing, we have the ability to buy exactly what we need for a given application set. The problem is knowing what you need. *ClusterWorld* believes that meeting this challenge will help grow the market and has therefore decided to take the lead in this effort.

To this end, *ClusterWorld* has recently begun the ClusterWorld Benchmark Project (CWBP). The goal of this project is to collect tools that can assist users and vendors in the process of evaluating cluster performance. Specifically, we envision several levels of tests that will help users make informed decisions and vendors build better products.

Inevitably, a benchmark implies a contest to see who has the fastest machine. These types of efforts have their own reward, but the goal of this project is not to boil results down to one number, but rather to provide a series of numbers that can be used to make better decisions about clusters.

Although there are currently no official guidelines, there are several tenets that will be used to help shape the project.

- For the most part, the benchmarks will be based on open source efforts. This requirement means that tests will be open and results will be easily reproducible.

- Benchmark suites and instructions will be freely available for download by users and vendors.
- The benchmarks will be easy to use and install. As an example, the Beowulf Performance Suite (www.hpc-design.com/download.html) is an RPM package that includes a basic GUI for running tests and provides results in html format.
- There will be no attempt to provide a single number as a means to rank clusters.

Intended Use

The primary use of benchmarking tools would be to assist in the evaluation phase of cluster acquisition. Ultimately the true test is your application(s), but in many cases testing this is not possible. For instance, if you know that most of your jobs will be bioinformatics based, then a group of tests that run real codes would be invaluable to your purchase decision. Another beneficial side effect is that by asking vendors to run tests, the customer attains a certain level of vendor confidence.

Another important issue relates to software and hardware upgrades. For instance, after installing a new version of the kernel, it is reasonable to ask, "Did this change hurt or help my performance." The same question can be asked about almost every upgrade or change made to the system.

Furthermore, measuring change implies that there is a baseline of some sort to compare against.

Without this type of informa-

tion, it is very difficult to know in what direction your cluster is headed when you change things.

With cluster computing, we have the ability to buy exactly what we need for a given application set.

Because users may need a spectrum of information, the CWBP will provide a hierarchy of tests starting off with low-level measurements and finishing with a workflow assessment.

In general, as one progresses through the benchmark hierarchy, the amount of run-time and user effort will increase. (For example, micro-benchmarks may take an hour to run, but workflow benchmarks may take several days.)

The following is a proposal for various layers of the benchmark suite.

I. Micro-Benchmarks

These types of benchmarks require one or two nodes and look at very low-level performance features, which may include memory bandwidth, multiprocessor scaling, network speeds, storage speeds, and other single-system tests.

II. Whole System Benchmarks

At a slightly higher level, a set of systemwide general benchmarks and tests will provide a sense of how the cluster is functioning as a whole system. Examples of these types of benchmarks include NAS Parallel Benchmarks, the HPL benchmark, and some of the MPI suites that are available.

III. Application Benchmarks

At this level, real applications will be tested. These will include representative applications from areas of bioinformatics, weather forecasting, computational chemistry, rendering, and others. As stated, these will, in part, be based on freely available applications.

IV. Workflow Benchmarks

The topmost level will test how well a specific workflow is handled by the system. This level obviously brings in the batch scheduling capabilities of the cluster. It requires a set of applications from level III that may look similar to the users' typical workflow. These applications would be queued on the cluster and then run to determine how much work gets done in a specific amount of time.

It is envisioned that the first version of the test suite will have a limited set of tests and a criteria for adding more tests. In this way, the infrastructure can be solidified prior to expanding the scope of the project.

Next Steps

Admittedly, at this point there are more questions than answers. Indeed, I would like to focus on two main questions:

1 Does this seem like a reasonable approach?

2. Who would like to help?

As a start, there will be a BOF (Birds-of-a-Feather) session at ClusterWorld Conference and Expo in April to discuss this project. We expect both users and vendors to attend this session.

In addition, this article will be posted on *www.ClusterWorld.com*

Example Benchmarks

To illustrate the kinds of benchmarks, some examples are given below.

Micro Benchmarks

Bonnie++ hard drive performance

- www.coker.com.au/bonnie++

Stream memory performance

- www.cs.virginia.edu/stream

Netperf general network performance

- www.netperf.org/netperf/NetperfPage.html

Netpipe detailed network performance

- www.scl.ameslab.gov/Projects/ClusterCookbook/nprun.html

Unixbench general Unix benchmarks

- www.linuxdoc.org/HOWTO/Benchmarking-HOWTO.html

LMbench low level benchmarks

- www.bitmover.com/lmbench

Whole System benchmarks:

NAS Parallel tests

- www.nas.nasa.gov/Software/NPB

HPL

- www.netlib.org/benchmark/hpl

MPI Tests

- icl.cs.utk.edu/projects/llcbench
- www-unix.mcs.anl.gov/mpi/mpich/download.html
- www.lam-mpi.org/7.0/test-suite.php

MPI Link-checker

- www.microway.com/mpilinkchecker.html

Application Benchmarks

NAMD

- www.ks.uiuc.edu/Research/namd

BLAST

- www.ncbi.nlm.nih.gov/Ftp

MPI-BLAST

- mpiblast.lanl.gov

GROMACS

- www.gromacs.org

MM5

- box.mmm.ucar.edu/mm5

WorkFlow Benchmarks

Grid Engine

- gridengine.sunsource.net

Torque

- www.supercluster.org/projects/torque

so readers can provide feedback and hold an online discussion. Of course you always can email me with your comments and suggestions (address below).

I also plan on putting together an advisory board to help guide the effort.

Next month, I should have

some of the issues nailed down and hopefully get some people committed to the advisory board so we can begin to move forward with the project.

Douglas Eadline, Ph.D. is the Editor-in-chief of ClusterWorld Magazine. He can be reached at deadline@clusterworld.com.