Yibin Dong, CBIL

Performance Measure of Microsoft Compute Cluster Server (CCS) 2003

June 3rd, 2006

Purpose:

The experiment was conducted to prove the concept of using distributed computing technologies for bioinformatics applications. The purpose was to show:

(a) Given a fixed complexity of an algorithm, the running time T will be decreased by a factor $\frac{1}{N}$, where N is the number of nodes in the cluster.

(b) Increasing number of nodes in the cluster will enable the cluster to handle jobs with higher complexity.

Demo Codes:

Demo codes are provided by the MathWorks with modification so that it can be run on Windows CCS. The demo uses the high-resolution ovarian cancer data set that was generated using the WCX2 protein array.
http://www.mathworks.com/products/demos/bioinfo/biodistcompdemo/biodistcompdemo.html

Dataset:

The dataset, High Resolution SELDI-TOF Study Sets, is provided by the FDA-NCI Clinical Proteomics Program Databank:
http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp
There are two classes, Ovarian cancer case(n=121) vs. high-risk control (n=95).

Experiments

There are two independent CCS clusters, which were built on two set of machines. The first cluster is a testing bed for CCS program debugging. It has two DELL XPS x64 machines (Intel Pentium 3.8GHz, 2G RAM). The second cluster is a live compute cluster with one head node server and six compute nodes (HP Proliant DL145 Generation 2 Server, dual core AMD Opteron processor 270, 2.01 GHz, 1G RAM).

A job with 216 (=121+95) tasks is created for testing. Each task takes a different input data, a sample profile, and generates an output. After all tasks are done, the head node collects the results from each compute node and saves in an output file. Figure 1 shows the running time of completing the job when varying the number of nodes in a cluster.

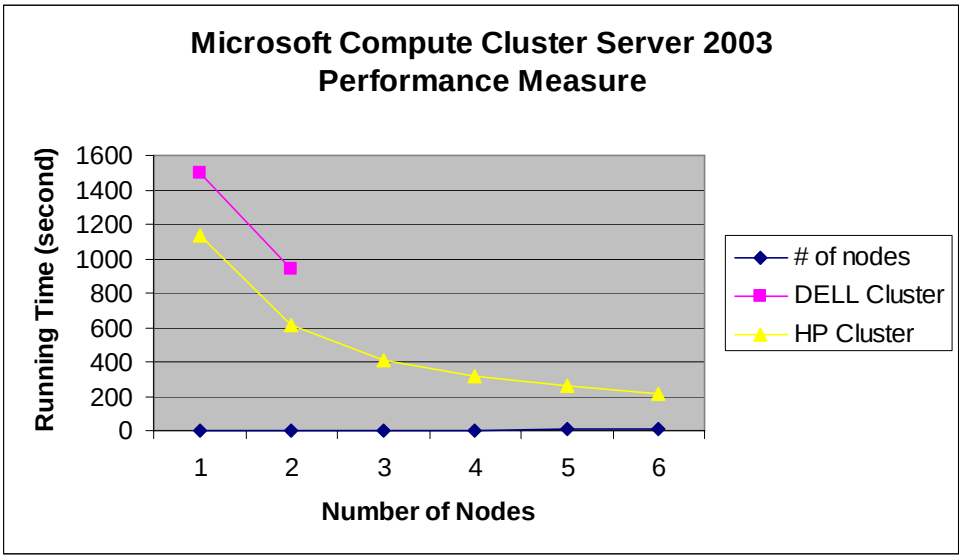

Figure 1. The running time to accomplish a job of 216 tasks when the number of compute nodes in a cluster is changing. The curve with square dots shows the performance measure of the DELL clusters with total 2 nodes. The curve with triangle dots shows the performance measure of the HP cluster with total 6 nodes available.