Affordable Supercomputing Solutions: Cluster Computers In Business Applications

Fred L. Kitchens\textsuperscript{a}, Sushil K. Sharma\textsuperscript{b}

\textsuperscript{a}Department of Information Systems and Operations Management
Ball State University
WB-203, Muncie, IN 47306, USA
765-285-5305 (office), 765-285-5308 (FAX)
Email: fkitchens@bsu.edu

\textsuperscript{b}Department of Information Systems and Operations Management
Ball State University
WB-203, Muncie, IN 47306, USA
765-285-5315 (office), 765-285-5308 (FAX)
Email: ssharma@bsu.edu

Abstract

Businesses need fast computing environments to quickly process the voluminous transactions and to respond to market requirements. Most companies cannot afford the cost of a supercomputer. As an experimental design, the authors have developed a cluster-computing project at their institution, to expose business students to the process of building and operating a cluster computer for use in supercomputing applications. The project helps students learn the technologies of distributed computing, parallel computing, networking, securing networked environments and designing applications for distributed computing. The purpose of this paper is to share the experiences of cluster computer implementation in a business curriculum and how cluster computing can be exploited for solving many complex real world business applications in a wide variety of industries.

Key Words:
Cluster Computing, Beowulf, course development, high performance computing, super computer

Introduction

Over the past decade, the computing industry has experienced a dramatic shift away from mainframe computing to a more distributed, client-server approach. In recent years, cluster computers have emerged as the leaders in high-performance computing (Sterling 2000). Today’s cluster-computers are more than a match for mainframes in terms of both cost and performance (Baker, Fox et al. 1995). Harnessing the combined computing power of multiple microprocessors in a parallel configuration. There are many cluster configurations, but a simple architecture such as the one shown in Figure 1, is used to visualize basic concept. In a typical cluster, the application is run on a Master node. However, the computational work is split-up and parsed out to be done by the multiple nodes in the cluster. This way, the cluster in better equipped to handle larger amounts of data and complex problems than a stand-alone PC.
A cluster computer has many advantages over traditional supercomputers in cost, performance, and maintenance (Supercomputing 2002). High-performance parallel computers incorporate proprietary interconnected networks allowing low-latency, high bandwidth and inter-processor communications. This has led to the recognition that cluster-computers are a realistic alternative for a variety of applications similar to that of the more expensive supercomputers (Baker, Fox et al. 1995). Many educational institutions and companies are beginning to explore this option as a cost-effective alternative for supercomputing applications, and retailers are looking to sell them these systems (Read 2002).

Cluster computing is a forward-looking process in high-performance computing. One of the important objectives of this type of computing system is to make super-computer-type processing speed available at an affordable price for users who previously could not meet the expense of a traditional super-computer. Cluster-computers consist of a dedicated network of personal computers, connected and operating in such a way as to harness the cumulative computing power of all computers in the cluster. The process involves using either new or old computers – even computers destined for the landfill – and through networking them together, harness their collective computing power (Kitchens and Sharma 2002). Cluster computers are expected to be the future of high performance computing. They offer a scalable, flexible, and reliable means of providing similar computing power to that found in traditional supercomputers (Combariza 2002). As technology advances, there is an increasing need for greater computing capacity within businesses and educational institutions alike. Applications in data mining, data
warehousing, simulations and forecasting, to name just a few, can utilize increased computing resources. At the same time, the value received from these practices is increasing. Advances in cluster computing will allow mid-size businesses, organizations, and institutions access to computing power previously unimaginable, at a fraction of the cost usually associated with supercomputing capacity (Supercomputing 2002). As an experimental design, the authors have developed a cluster-computing project at their institution, to expose business students to the process of building and operating a cluster computer for use in supercomputing applications. The project helps students to learn the technologies of distributed computing, parallel computing, networking, securing networked environments and designing an application for distributed computing. The purpose of this paper is to share the experiences of cluster computer implementation in a business curriculum and how Cluster Computing can be exploited for solving many complex real world business applications in a wide variety of industries.

**Beowulf Cluster Computing Project – A Successful Experiment**

Cluster computing has emerged as an opportunity for businesses and academic institutions to garner supercomputing-processing power in their respective practices (Sterling 2000). Building a cluster computer is an alternative to purchasing an expensive supercomputer. Breaking from similar projects at other institutions where cluster computing systems have been built by purchasing brand new equipment in attempt to construct the most powerful system possible, the Cluster Computing project at author’s institution is aimed at building a system with the highest performance possible for the least investment possible. Nearly all equipment used in the project was acquired from the university’s excess inventory or donated from local businesses and private citizens from the surrounding community.

The Authors’ Cluster Computing Research Project began as a project with little background or direction as far as previous course development projects to model from other academic institutions. The project has become successful, as the group learned a great deal through trial-and-error. Students have learned not only how to build a cluster but also the application of various concepts of MIS courses (Apon, Buyya et al. 2001). Presently, the authors are operating a cluster-computer consisting of 23 nodes. All of the equipment comes from used and excess inventory of the university and local business. If not for the authors’ cluster computing project, the equipment currently in use would have been sold at auction, recycled, or donated to charity (Childers, Kitchens et al. 2003) (Childers, Kitchens et al. 2003) (Kitchens and Sharma 2002). Working around certain bottlenecks, a cluster computer increases in speed and capability as individual nodes are added, or exchanged for improved nodes. The potential power that becomes available through cluster computing is demonstrated in Figure 2. The cluster is used to provide hands-on experience on a variety of practical applications in a number of courses such as hardware, operating systems, networking systems, security, system analysis and design and project management (Childers, Kitchens et al. 2003). Students built the cluster themselves. By doing so, they gained an integrated and hands-on experience in multiple concepts that are otherwise covered through various courses (Childers, Kitchens et al. 2003).
Having established a basis for cluster computing in the curriculum, the authors are interested in expanding the scope of the project. The objective is to bring supercomputing to business. Most supercomputing applications are currently developed for scientific research. With an affordable supercomputing option now available, the authors seek to develop business applications in areas where none currently exist. A multitude of applications and ideas are waiting to be discovered and developed. Several potential applications for business are identified and discussed in the next section.

**Cluster Computing for Business Applications**

Clustered computers are specifically designed to take large programs and sets of data, subdivide them into component parts, thereby allowing the individual nodes of the Cluster to process their own individual “chunks” of the program. The sophisticated messaging and control protocol of the Cluster allows this to be done simultaneously, even if the various nodes can handle their “chunks” at different rates.

Most cluster computing applications to date have been in scientific fields, such as astronomy, physics, chemistry, biology, etc… The authors believe the field of business represents a vast untapped field of applications for cluster computers. One of many methods of introducing
Cluster computing into business fields is to teach it in college and allow graduates to infiltrate the business world, carrying their knowledge with them. Traditionally, in the few cases where cluster computing has been taught in college, it has been in Computer Science or Engineering. The author’s course is the first cluster-computing course in any college of business in the world.

Cluster computing has proven to be an extremely powerful tool to be used for a variety of multidisciplinary teaching applications. These applications have provided students a unique learning experience. Applications in courses such as Networking, Programming, MIS, Security, Statistics, Artificial Intelligence, Database, and Systems Analysis and Design have utilized the computing power of cluster computing. The use of cluster computing in these courses provided students with opportunities previously unavailable. With the experience of developing a cluster-computer and utilizing the parallel environment to successfully teach various MIS courses, the authors are experimenting with cluster computing for various business applications in a variety of industries as shown in Figure 3. The author’s Cluster Computing course represents a great hands-on capstone course integrating many aspects of the IS curriculum. Potential business applications are discussed and students are encouraged to brainstorm additional applications. Students are prepared for a future in which cluster computers will be the norm in business.

**Economics: “synthetic economies”**

A potential business application of cluster computing is in economic modeling, more specifically in the field of consumer behavior. The new supercomputing environment will allow researchers to explore theories in “synthetic environments,” or environments that have received their variable attributes from other research and data gathering methods. Once the parameters for these synthetic environments have been established, researchers can perform calculations to determine how elements will react (or interact) within this environment. The exploration of “synthetic economies” through simulations could help to predict what consumers will do under certain circumstances. In addition, the economic modeling will help identify consequences of certain consumer behaviors. Some possible economic applications include investigating how
consumers will respond to new promotional campaigns, consumer reactions to product price changes, changes in consumer behavior with respect to newly introduced products, the effects of new companies entering/exiting industries, and the possible effects of regulatory and political decisions (Purdue 2002). From a programming standpoint, these applications could take the form of neural networks, Monte Carlo simulations, or even interactive modeling; all of which are well suited to cluster computing environments.

Real Estate Industry: Brownfields Redevelopment

Another business application of cluster computers is in the real estate industry, especially in the context of “Brownfields” redevelopment projects. The U.S. Environmental Protection Agency (EPA) defines a Brownfield as a “real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant” (Agency 2002). Identifying potential Brownfield sites is an important activity in the real estate industry. One of the methods used to identify Brownfields is historical data analysis, done with the aid of GIS (Geographic Information Systems). In addition to the speed efficiencies, a cluster computer would do a more extensive site analysis over a greater geographical area with more accuracy. Analyses such as this can be greatly enhanced through the use of cluster computers, largely because the increased computational ability of a cluster can allow more variables and more data to be processed in a shorter time. This can produce improved results in less time.

Public Administration: Crime Analysis

Cluster computing applications have great potential in in public administration. This is one of the most beneficial applications of cluster computer systems. In order to perform such analysis and modeling, it would be necessary to create a framework very similar to the “synthetic environment” discussed in the economics application. Necessary elements include: 1) a method to define the entities and their interactions/behaviors, 2) automated evolution of a population and background entities, 3) time management techniques, and 4) An appropriate model of “the environment” and interactions of entities within the environment (Gottschalk 1999). The cluster can then use these data sources to build relatively accurate environments representing the crime conditions for a specified geographic area. As users interact with this environment and alter certain variables, the crime conditions will then reflect those changes through further analysis and modeling. The result will be a changed crime environment, possibly showing where crime trends are most likely to shift next. The future shift in crime can be predicted based upon the risk factors previously discussed. An application such as this, requiring spatial and geographical representations and as many variables as possible, are ideally suited to the unique data-handling and number crunching nature of cluster computers.

Environmental Industry: Ecosystems

Clusters could be used to, “model large numbers of individual, locally interacting organisms, and explore both the characteristics of equilibrium and the consequences of disrupting this balance” (Gottschalk 1999). This is an accurate description of a basic ecosystem. Furthermore, this definition of an ecosystem lends itself toward analysis through cluster computing. There are many reasons that individuals or organizations would want to model interactions with the ecosystem. One use would be to determine the positive or negative consequences of a large-
scale development. Other reasons would include attempting to determine what population level the Earth is capable of supporting at the equilibrium point, or determining the effects of historic, current, and future pollution controls. Applications such as this involve the monitoring of specific virtual entities, each possessing unique characteristics and interaction behavior. Studying the simultaneous actions and interactions of multiple entities is an application that can best be handled by cluster computers.

**Manufacturing Industry**

One method in which the manufacturing industry can benefit from cluster computing applications is to improve the efficiency of operations on the shop floor. Individual nodes of the cluster can collect information about the processes at specific points and times on the shop floor. Then, the cluster can process the data to produce models that will root-out efficiencies and produce schedules for production (Kitchens and Sharma 2003). While this type of analysis may be similar to an ERP system, a cluster computer is another step in the evolution of ERP. As a real-time system, it can be made to interact with the users – scheduling breaks, down time, and retooling periods as well as facilitating assembly-line shutdowns to allow partial- rather than full-shutdowns depending on the particular incident, production requirements, and bottlenecks.

**Artificial Neural Networks**

A practical application for cluster computer systems can be found in artificial neural networks. Cluster computers can be used to reduce the time required to train artificial neural networks; one of the most common complaints of this technique. This type of application is particularly beneficial in several key industries, but it certainly has applications in many fields. Artificial neural networks have, for example, been used in the financial industry, oil and natural gas industry, real estate industry, credit industry, medical industry, mental health industry, mining industry, manufacturing industry, and public administration, just to name a few. Two ideal applications for artificial neural networks trained on cluster computers are real estate and advertising.

**Real Estate Industry: Appraisal Values:**

Another application for Beowulf clusters in the real estate industry is predicting appraisal values. The real estate market consists primarily of buyers and sellers, attempting to find housing that meets a wide variety of needs, including price. The appraisal process is the primary process used to determine a home’s “market value.” However, the problems inherent with these valuation methods are the inconsistency between appraisers, the inability of machines to consider more than rules and mathematical formulas, and the effects of changing market conditions. Because neural networks are objective, unbiased, and able to continually learn, they provide an ideal solution to the appraisal problem. Neural networks are ideally suited to run on a cluster-computing platform; capable of using more variables and more data, producing more accurate results in less time.

**Advertising Industry: Direct Mail Users:**

Another area that can potentially benefit from the use of neural networks trained on cluster computers is in the advertising industry, specifically, anywhere that direct mail advertising is
employed. Artificial neural networks have been used to target increased response rates through direct mail. Since the purpose of the direct mail effort is to generate revenue, it is important to seek out those that are most likely to generate income (BrainMaker 2002). Cluster computers will allow faster analysis, using more variables and more data, producing improved analysis.

The authors have identified only a few business applications for cluster computers. However, many other potential applications exist, in a wide variety of industries.

**Conclusion**

This paper has demonstrated just a few of the many practical applications for cluster computer in business. The wide variety of applications that exist throughout many different industries also presents cluster computer systems as a versatile solution, one that is certainly not designed for niche markets. Although not every business will have a feasible application for a Beowulf cluster, it is evident that the idea of a Beowulf cluster no longer belongs solely to the realm of science..

The authors believe that the time has come for cluster computers to enter the business realm and that one method of introduction is through university business curricula. In the author’s experience, cluster computing provides a great hands-on integrated experience for students. It can easily be used in a capstone course, and it provides students an introduction to the next great paradigm in business technology.

**References**


