International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 2, Number 3 (2012), pp. 133-140

http://www. ripublication.com

Energy Efficient Resource Management using Green Virtualisation Cluster Computing

K.Sriprasadh[1], S.Magesh kumar[2] , K.Mohan[3], M.Gopinath[4]

Department of Computer Science and Engineering
Department of Information Technology
Thirumalai Engineering College, Kanchipuram
E-mail: srisaiprasadhhh@gmail.com[1], mageshkumars@yahoo.com[2],
jackmoh2000.2009@gmail.com[3], gopinath_it2005@yahoo.com

Abstract

The emerging technique for environment friendly computing is by clustering the multiple servers into a single server this makes power consumption and resource segregation at a single point this enables a multiple users can access multiple server in single server mode. This provides cost consumption and energy consumption that will provide a green computing environment.

Keywords: Virtualization, Cluster Computing, Energy Consumption.

Introduction

Normally Cluster computing are exhibiting the greatest rate in growth of any class of parallel computer and may dominate high performance computing in the near future Parallel processing and parallel computer architecture is a field with decades of experience that clearly demonstrates the critical factors of interconnect latency and bandwidth, the value of shared memory, and the need for lightweight control software. Generally, clusters are known to be weak on all these points. Bandwidths and latencies both could differ by two orders of magnitude (or more) between tightly couple MPPs and PC clusters. The shared memory model is more closely related to how applications programmers consider their variable name space and such hardware support can provide more efficient mechanisms for such critical functions as global synchronization and automatic cache coherency. And custom node software agents can consume much less memory and respond far more quickly than full-scale standalone operating systems usually found on PC clusters. In fact, for some application classes these differences make clusters unsuitable. But experience over the last few years has shown that the space and requirements of applications are rich and varying. While some types of applications may be difficult to efficiently port to clusters, a much broader range of Workloads can be adequately supported on such systems, perhaps with some initial effort in optimization. Where conventional application MPP codes do not work well on clusters, new algorithmic techniques that are latency tolerant have been devised in some cases to overcome the inherent deficiencies of clusters. As a consequence, on a per node basis in many instances applications are performed at approximately the same throughput as on an MPP for a fraction of the cost. Indeed, the price-performance advantage in many cases exceeds an order of magnitude. It is this factor of ten that is driving the cluster revolution in high performance computing.

A "*commodity cluster*" is a local computing system comprising a set of independent computers and a network interconnecting them. A cluster is *local* in that all of its component subsystems are supervised within a single administrative domain, usually residing in a single room and managed as a single computer system. The constituent *computer nodes* are commercial-off-the-shelf (COTS), are capable of full independent operation as is, and are of a type ordinarily employed individually for standalone mainstream workloads and applications. The nodes may incorporate a single microprocessor or multiple microprocessors in a symmetric multiprocessor (SMP) configuration. The *interconnection network* employs COTS local area network (LAN) or systems area network (SAN) technology that may be a hierarchy of or multiple separate network structures. A cluster network is dedicated to the integration of the cluster compute nodes and is separate from the cluster's external (worldly)environment.

A cluster may be employed in many modes including but not limited to: high capability or sustained performance on a single problem, high capacity or throughput on a job or process workload, high availability through redundancy of nodes, or high bandwidth through multiplicity of disks and disk access or I/O channels. A "*Beowulf-class system*" is a cluster with nodes that are personal computers (PC) or small symmetric multiprocessors (SMP) of PCs integrated by COTS local area networks (LAN) or system area networks (SAN), and hosting an open source Unix-like node operating system. An Windows-Beowulf system also exploits low cost mass market PC hardware but instead of hosting an open source Unix like O/S, it runs the mass market widely distributed Microsoft Windows and NT operating systems. A "*Constellation*" differs from a commodity cluster in that the number of processors in its node SMPs exceeds the number of SMPs comprising the system and the integrating network interconnecting the SMP nodes may be of custom technology and design.

This technique is used in computing as a multiple user are tend to use a single data at a same time, so wastage of waiting time makes loss of energy, This heats up the system and causes excretion of system heat this increases environment temperature ,this can be avoided by new green computing methodology so called as Green Clustering Computing.

Benefits of Clustering Computing

Clusters allow *trickle-up*; hardware and software technologies that were developed for broad application to mainstream commercial and consumer markets can also serve in the arena of high performance computing. It was this aspect of clusters that initially made them possible and triggered the first wave of activity in the field. Both network of workstations and Beowulf-class PC clusters were possible because they required no expensive or long-term Development projects prior to their initial end use. Such early systems were far from perfect but they were usable. Even these inchoate cluster systems exhibited price-performance advantage with respect to contemporary supercomputers that approached a factor of 50 in special cases while delivering per node sustained performance for real-world applications often within a range of a factor of 3 and sometimes well within 50% of the more costly systems with the same number of processors. But the rapid rate of improvement in PC microprocessor performance and advances in local area networks have led to systems capable of tens or even hundreds of Gigaflops performance while retaining exceptional price performance Commodity clusters permit a flexibility of configuration not ordinarily encountered through conventional MPP systems. Number of nodes, memory capacity per node, number of processors per node, and interconnect topology are all parameters of system structure that may be specified in fine detail on a per system basis without incurring additional cost due to custom configurability. Further, system structure may easily be modified or augmented over time as need and opportunity dictates without the loss of prior investment. This expanded control over system structure not only benefits the end user but the system vendor as well, yielding a wide array of system capabilities and cost tradeoffs to better meet user demands. Commodity clusters also permit rapid response to technology improvements. As new devices including processors, memory, disks, and networks become available, they are most likely to be integrated in to desktop or server nodes most quickly allowing clusters to be the first class of parallel systems to benefit from such advances. The same is true of benefits incurred through constantly improving price-performance trends in delivered technology. Commodity clusters are best able to track technology improvements and respond most rapidly to new component offerings.

Cluster Network Hardware

Commodity clusters are made possible only because of the availability of adequate inter-node communication network technology. Interconnect networks enable data packets to be transferred between logical elements distributed among a set of separate processor nodes within a cluster through a combination of hardware and software support. Commodity Clusters incorporate one or more dedicated networks to support message packet communication within the distributed system. This distinguishes it from ensembles of standalone systems loosely connected by shared local area networks (LAN) that are employed primarily as desktop and server systems. Such computing environments have been successfully employed to perform combined computations using available unused resources.

These practices are referred to as “cycle harvesting” or “workstation farms” and share the intercommunication network with external systems and services, not directly related to the coordinated multi-node computation. In comparison, the commodity cluster's system area network (SAN) is committed to the support of such distributed computation on the cluster, employing separate external networks for interaction with environment services.

Servers Are Driving Energy Consumption and Costs

Today's datacenters consume a lot of electricity. A recent report by the Environmental Protection Agency claims datacenters in the U.S. consume 4.5 billion kWh annually, 1.5 percent of the country's total .Perhaps more importantly; this figure has doubled from 2000 to2006, and is likely to double again in the next few years. This trend is affecting datacenters around the world and is likely to continue, given how central computing is to our businesses and lifestyles. There are many factors contributing to excessive energy consumption in datacenters, but underutilized x86 hardware is the most significant. According to the EPA, servers consumed 80 percent of the total IT load and 40% of total datacenter power consumption.

Site infrastructure—including cooling of equipment—accounts for another 50 percent of total datacenter power consumption. Yet because x86 servers typically house only a single application, their processors sit idle 85-95 percent of the time. While sitting idle, these servers use nearly as much power as they do when they are active. According to analysts, companies maintain roughly three years of excess hardware capacity due to this vast underutilization. With more than seven million servers sold annually, this represents more than 20 million servers sitting idle and wasting energy. This inefficiency is not only wasteful but expensive, especially as electricity costs and computing demand continue to rise.

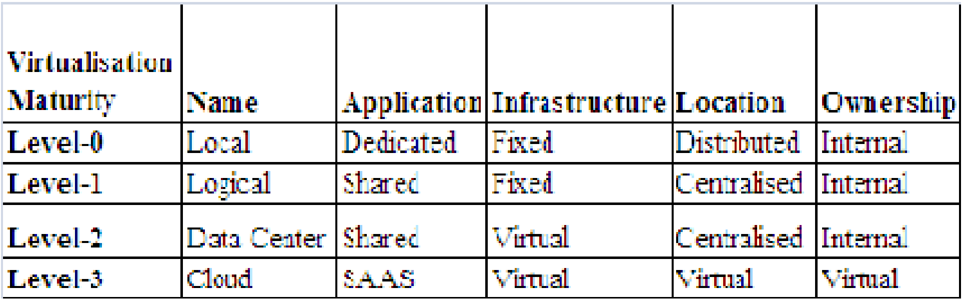
As a result of increasing energy demands on inefficient and aging datacenters, many companies are simply running out of power and /or capacity.[2] Either the utility cannot provide adequate power, or the equipment is so power-hungry and dense that the datacenter runs out of capacity even though the datacenter is not physically full; energy costs preclude investments in additional physical hardware. Analyst firms and industry research suggest that most datacenters will feel the crunch soon, if they don't already. To overcome from this we introduce a new concept of clustering the servers with virtualization in that a new virtualization technique is included.

Server Consolidation

A key benefit of virtualization technology is the ability to contain and consolidate the number of servers in a datacenter. This allows businesses to run multiple application and OS workloads on the same server. Ten server workloads running on a single physical server is typical, but some companies are consolidating as many as 30 or 40 workloads onto one server. As you might expect, dramatically reducing server count has a transformational impact on IT energy consumption. Utilization of x86 servers increases from the typical 8-15 percent to 70-80 percent. Reducing the number of physical servers through virtualization cuts power and cooling costs and provides more computing power in less space. As a result, energy consumption typically decreases by 80 percent The impact of virtualization on energy consumption is so significant that utilities in North America such as PG&E, Southern California Edison, SDG&E, BC Hydro and Austin Energy are paying customers for removing servers through consolidation.iii These programs compare the energy use of existing equipment to that of remaining equipment in service after consolidation. Incentives are based on the net reduction in kilowatt-hours from direct energy savings from the project (cooling costs are excluded), which can be as high as $300 USD per server and $4 million per physical site. Incentive programs are more cost effective than creating new power plants, and better for the environment.

Figure 1: Virtualized server

DRS use a feature called VMware V Sphere Distributed Power Management (DPM) to reduce power consumption by turning off servers when there is unneeded capacity. Servers are powered back on when the capacity is required. Because virtual machines are unaffected by live migration, this feature automatically shrinks or expands the pool of servers running at any given time without reducing service levels. This capacity on demand eliminates the need to maintain “excess capacity” while ensuring resources are available if more capacity is needed. DRS also reserves capacity for automatic failover Virtualization is a term used to mean many things, but in its broader sense, it refers to the idea of sharing. To understand the different forms of virtualization and the architectural implications for creating and deploying new applications, we propose a reference model to describe the differing forms of the concept.



Virtualisation Maturity	Name	Application	Infrastructure	Location	Ownership
Level-0	Local	Dedicated	Fixed	Distributed	Internal
Level-1	Logical	Shared	Fixed	Centralised	Internal
Level-2	Data Center	Shared	Virtual	Centralised	Internal
Level-3	Cloud	SAAS	Virtual	Virtual	Virtual

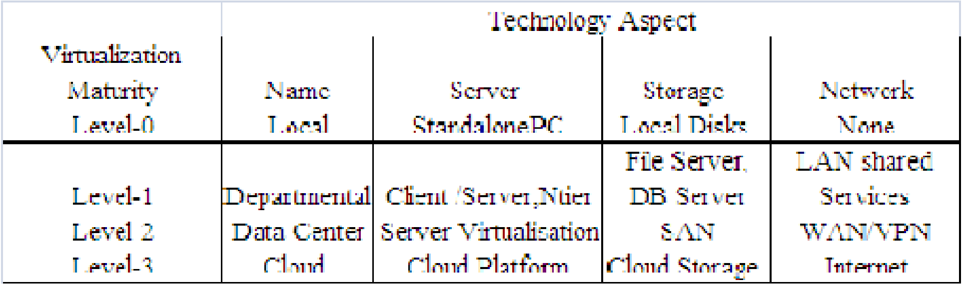
In this model we observe a number of different layers of abstraction at which virtualization can be applied, which we describe as increasing levels of maturity, shown in Table 1.We assert that higher levels of virtualization maturity correspond to lower [3] energy consumption, and therefore architectures based on higher levels of maturity are “greener” than those at lower levels, which we discuss further on.

Level 0 (“Local”) means no virtualization at all. Applications are all resident on individual PCs, with no sharing of data or server resources.

Level 1 (“Logical Virtualization”) introduces the idea of sharing applications. This might be, for example, through the use of departmental servers running applications that are accessed by many client PCs. This first appeared in the mainstream as mainframe and then “client/server” technology, and later with more sophisticated N-tier structures. Although not conventionally considered virtualization, in fact, it is arguably the most important step. Large organizations typically have a large portfolio of applications, with considerable functional overlaps between applications. For example, there may be numerous systems carrying out customer relationship management (CRM) functions.

Level 2 (“Data Center Virtualization”) is concerned with virtualization of hardware and software infrastructure. The basic premise here is that individual server deployments do not need to consume the hardware resources of dedicated hardware, and these resources can therefore be shared across multiple logical servers. This is the level most often associated with the term virtualization. The difference from Level 1 is that the hardware and software infrastructure upon which applications/servers are run is itself shared (virtualized). For server infrastructure, this is accomplished with platforms such as Microsoft Virtual Server and VMware among others, where a single physical server can run many virtual servers. For storage solutions, this level is accomplished with Storage Area Network (SAN) related technologies, where physical storage devices can be aggregated and partitioned into logical storage that appears to servers as dedicated storage but can be managed much more efficiently. The analogous concept in networking at this level is the Virtual Private Network (VPN) where shared networks are configured to present a logical private and secure network much more efficiently than if a dedicated network were to be set up.

Level 3 ("Cloud virtualization") in the virtualization maturity model extends Level 2 by vitalizing not just resources but also the location and ownership of the infrastructure through the use of cloud computing. This means the virtual infrastructure is no longer tied to a physical location, and can potentially be moved or reconfigured to any location, both within or outside the consumer's network or administrative domain. The implication of cloud computing is that data center capabilities can be aggregated at a scale not possible for a single organization, and located at sites more advantageous (from an energy point of view, for example) than may be available to a single organization. This creates the potential for significantly improved efficiency by leveraging the economies of scale associated with large numbers of organizations sharing the same infrastructure. Servers and storage virtualized to this level are generally referred to as Cloud Platform and Cloud Storage, with examples being Google App Engine, Amazon Elastic Compute Cloud, and Microsoft's Windows Azure. Accessing this infrastructure is normally done over the Internet with secure sessions, which can be thought of as a kind of virtualized discrete VPN. Each level of maturity has a number of significant technologies "aspects" of the computing platform that may be virtualized. A summary of the virtualization layers as they map to the server, storage, and network aspects is shown in Table 2.



Virtualization Maturity	Technology Aspect			
	Name	Server	Storage	Network
Level-0	Local	StandalonePC	Local Disks	None
Level-1	Departmental	Client /Server,N-tier	File Server, DB Server	LAN shared Services
Level 2	Data Center	Server Virtualization	SAN	WAN/VPN
Level-3	Cloud	Cloud Platform	Cloud Storage	Internet

Energy Costs of a New Virtualization Server

We define the following:

- N – The number of servers to be virtualized on a single new physical server
- B – Embodied energy ratio (embodied energy of new server divided by total energy consumption of that server over its life cycle)
- E – Efficiency factor (energy consumption of a single new server with capacity equivalent to the original N servers divided by energy consumption of N original servers, assuming the same technology and utilization, for the projected life) T – Technology factor (energy consumption of new servers per unit
- CPU capacity divided by energy consumption of old servers per unit CPU capacity) U = utilization factor (utilization of old servers divided by utilization of new server) to pay back the cost of embodied energy and realize a net gain, you need: $E \times U \times T < (1 - B)$

If a typical B value is 25 percent, then total improvement factors needs to be better than 0.75. This is easy to achieve since even if the technologies of old and new servers are similar (T= 1) and there is no efficiency gains (E=1) you would still expect U to be lower than 0.5 if N is greater than 2 since nearly all servers are grossly underutilized.

Conclusion

There is a compelling need for applications to take environmental factors into account in their design, driven by the need to align with organizational environmental policies, reduce power and infrastructure costs and to reduce current or future carbon costs. The potential reduction in energy and emissions footprint through good architectural design is significant. The move to more environmentally sustainable applications impacts software and infrastructure architecture. The link between the two is strong, driving a need for joint management of this area of concern from infrastructure and software architects within organizations.

References

[1] Cluster Computing White Paper 1 Final Release Version 2.0Editor - Mark Baker, University of Portsmouth, UK Propulsion Laboratory, USA
[2] How VMware Virtualization Right-sizes IT Infrastructure to Reduce Power Consumption at http://www.vmware.com/go/patents.
[3] The Architecture Journal, Green Maturity Model for Virtualization by Kevin Francis and Peter Richardson.