Study of Virtualization Energy-efficiency in High-energy Physics Computing

Jukka Kommeri[1], Tapio Niemi[1] and Marko Niinimaki[2]
[1]*Helsinki Institute of Physics, CERN, Geneva, Switzerland*
[2]*Hepia, Univ. Appl. Sci. West Switzerland, Geneva, Switzerland*

Keywords: High-energy physics, Virtualization, Energy-efficiency.

Abstract: Modern multi-core servers are able to run growing amount of physics analysis tasks. As the count of CPU cores keep growing, a need for sharing a single server among several analysis tasks becomes more difficult. The need for varying analysis environments increase the complexity of an computing node software collection. In this paper we study how virtual machines should be deployed and loaded when running high-energy physics analysis applications to achieve high throughput and minimal energy consumption. We build a test environment using a realistic data analysis software and performed a large set of test runs. We used both 4 core single processor and two processor 12 core servers to evaluate bottlenecks of physics analysis software. Our results indicate that both throughput and energy efficiency strongly depend on how many virtual machines (VM) are run in a computing node and how many analysis applications are processed in parallel in a VM: It is more efficient to have less VMs with more parallel applications than one application in each VM. Thus, we suggest that jobs of the same user running in the same environment should be combined to the same VMs instead of running each job in a different VM.

1 INTRODUCTION

Scientific computing clusters at CERN have traditionally allocated resources for one analysis job such that it gets one core and 2 GB of memory. As the number of cores in a CPU and the number of CPUs in the server increase, more jobs must be processed in parallel in the server. Modern servers can have 16 cores per CPU and hundreds of gigabytes of memory. Combining this with the need for different analysis environments, computing resources should be divided into smaller logical units. This is where virtualization comes in. Virtualization makes it possible to create logical containers, virtual machines, that contain a complete operating system with a user specific analysis environment. These can be modified to meet users resource requirements and moved from physical machine to another to improve the total energy efficiency larger server cluster.

In this study, we focus on studying how scheduling jobs in different ways among virtual machines affect energy consumption. We aim at providing how administrators should configure computing clusters to decrease energy consumption and improve throughput. Virtualization in this paper's context refers to system virtualization where several operating systems are run on single physical hardware.

We studied how computing jobs are distributed among users in CERN computing cluster as shown in Figure 1. According to the data more than 90% of users send more than one job and the majority of users send around 40 jobs in an hour. Based on this we can conclude that providing the user with one or more virtual machines and running several parallel jobs in each of them would be a working scheduling method. Thus, our hypothesis is that this kind of scheduling would offer better energy efficiency and total throughput than deploying an individual virtual machine for every job. Since virtual machines are not divided among users, privacy issues will not arise.

We use a realistic test environment and run several tests variating the number virtual machines and their load. The used virtualization system is the KVM open source hypervisor. We used a synthetic benchmark application, Lapack, and realistic HEP analysis jobs in the CMS software framework (CMMSW). The test environment consists of modern single processor and dual processor servers. In this way we can show that the results are not hardware depended. Finally, the measurements on virtual machines are compared to

Niemi T., Niinimaki M. and Kommeri J. (2012).
Study of Virtualization Energy-efficiency in High-energy Physics Computing.
In *Proceedings of the Sixth International Symposium on e-Health Services and Technologies and the Third International Conference on Green IT Solutions*, pages 77-82
DOI: 10.5220/0004474800770082

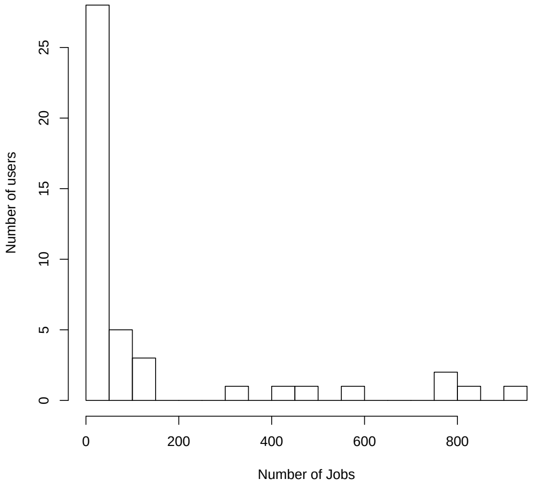

Figure 1: Job distribution per users for one hour.

those on non virtualized hardware. In all the tests, we measure energy consumption and processing time.

We found out that virtual machine configuration has a significant influence on the energy efficiency of the system. Also, the application that is run inside the virtual machine has an influence on how the virtual machines should be deployed on a physical hardware. In pure CPU load the overhead of virtualization is not as high as when running physics analysis work. With the physics analysis energy efficiency can be increased by decreasing the number of virtual machines and increasing their load.

2 RELATED WORK

Virtualization technologies are a key component of cloud computing (Buyya et al., 2008). Large data centers host cloud applications on thousands of servers (Schäppi et al., 2007; STAR, 2007). In such environments, the benefits of virtualization are obvious. Xu et al. (Xu et al., 2004) mention just-in-time compute and storage capacity, reducing management and administration cost through automation and providing greater control over end-user service levels.

Virtualization of the HEP grid clusters is not a new idea and has been studied by many researchers. Fenn et al. (Fenn et al., 2009) have tested high performance applications (HPC) in clusters that are made of virtual machines. They found KVM to be usable in non I/O intensive loads. Since those test there has been improvements to the KVM I/O and nowadays there is a paravirtualized drivers for KVM network and disk.

Regola et al. have studied the use of virtualization in high performance computing (HPC) (Regola and Ducom, 2010). They concluded that the I/O performance of full virtualization or para-virtualization is not yet good enough for low latency and high throughput applications such as Message Passing Interface (MPI) applications.

Virtualization technologies fall into categories of full virtualization and paravirtualization. As stated by Chaudhaury et al (Chaudhary et al., 2008), in full virtualization an unmodied operating system runs using a hypervisor to trap and safely translate/execute privileged instructions on-the-y. Paravirtualization, on the other hand, requires changes to the virtualized operating system.

Nussbaum et al. (Nussbaum et al., 2009) evaluated the full-virtualization and paravirtualization technologies with a cluster of 32 servers using a HPC Challenge benchmarks. They found the performance of full virtualization is far behind that of paravirtulization. Also sharing workload among different number of virtual machines did not seem make much difference. Verma et al. (Verma et al., 2008) also studied the effect of sharing same workload among different number of virtual machines. In their tests virtualization and division of load between several virtual machines did have much impact on overhead.

Padala et al. (Padala et al., 2007) have also studied performance of virtualization. They studied the effect of server load and virtual machine count on multi tier application performance. They found OS virtualization to perform much better than paravirtualization. The overhead of paravirtualization is explained by L2 cache misses; in the case of paravirtualization they increased more rapidly when the load increased.

Virtualization and energy consumption has been a subject of our earlier study (Kommeri et al., 2012) in which we compared KVM and Xen with a set of benchmarks. We found that idle consumption of Xen under Linux 3.0 is almost twice that of hardware, whereas KVM's idle consumption is only slightly higher than hardware's. Xen under Linux 2.6 however is closer to KVM. The mean power consumption of a server system with 1 virtual machine was about 110% of hardware's power consumption when using KVM and Xen.

3 METHODOLOGY

We run our test applications in a virtualized environment with varying parallelism both in the number of virtual machines on a hypervisor and the actual load inside a single virtual machine. Our hypothesis is that the system runs more energy-efficiently if we decrease the number of virtual machines and increase the load inside one virtual machine. We run our tests on several different hardware and measured processing times and energy consumption. As a comparison hardware version of the same measurements were made. In the following section, processing a job mean a single execution of any of the test applications in subsection 3.2.

3.1 Test Environment

In our test we used following hardware:

- 2CPU 12 core server, Opteron 2427, 32GB 800MHz memory, 1TB hard disk
- Dell Poweredge R210 II, Xeon E31260L, 8GB 1333MHz DDR3, 1TB hard disk (energy-efficient)
- Dell Poweredge R210 II, Xeon E31280, 8GB 1333MHz DDR3, 1TB hard disk (powerful)
- Dell Poweredge R210, Xeon X3430, 8GB 1333MHz DDR3, 250GB hard disk

Single CPU part of the tests in subsection 4.2 were done using a Dell R210 with Intel Xeon X3430. Results of these tests are illustrated by Figures 4 and 5.

As an operating system in all our servers and client machines we used default 64-bit Ubuntu LTS 10.04 with Linux kernel 2.6.32-40. Our virtualization environment consisted of default KVM hypervisor that run virtual machines using raw image files without paravirtualization. KVM is a Linux module that shows and schedules virtual machines as normal processes, which are scheduled by Linux scheduler (Kivity et al., 2007). Virtual machines were created with a distribution tool virt-install. Operating system in the virtual machines was Scientific Linux CERN 5 (SLC5) with Linux kernel 2.6.18-274.7.1.el5.

To study the effect of parallelism and virtual machine count on the energy efficiency, we used different virtual machine configurations. Tables 1 and 2 illustrate the configurations used in our hypervisors. 500 MB of memory was left for the host operating system. Otherwise memory and processor resources were divided equally among the virtual machines.

Power usage data was collected with a Watts up? PRO meter via a USB cable. Power usage values were recorded every second from the server hosting the virtual machines.

3.2 Test Software

The software used for our energy-efficiency tests consisted of Lapack 3.4.1, Linear Algebra PACKage, benchmark and CMSSW 4.2.4. Lapack is a collection of mathematical equations that are used to benchmark processors and extends Linpack benchmark that is used to benchmark Top500 servers (Anderson et al., 1990). Its execution time is less than one minute. To get more usable energy measurements the same program was run 10 times in a loop. This loop is considered as a job in the following sections.

Table 1: Settings for a single virtual machine in 12-core Opteron server.

VM count	VCPUs	Memory (GB)
1	12	31.5
4	3	7.88
8	2	3.94
16	1	1.97

Table 2: Settings for a single virtual machine in 4-core Dell 210 II.

VM count	VCPUs	Memory (GB)
1	8	7.5
2	4	3.75
3	3	2.50
4	2	1.88
5	2	1.50
6	1	1.25

CMSSW framework is used to analyse the data from LHC and (Fabozzi et al., 2008) This analysis task is a very typical one in high-energy physics. We used real data created at CERN. The data was stored in a ROOT image (Antcheva and et al., 2009) files, which our case were of size 4GB. Normally, a data analysis with this data can take days to perform. We limited the number of events of one analysis task to 300 events. With this limitation the analysis takes 10 minutes on the Opteron hardware. The data ROOT images were located on network file system, NFS, which was shared by all analysis tasks. The network traffic caused by one analysis task is very small, 2kB per task. The NFS server runs on a Dell T710 server and is connected to the same gigabyte ethernet local area network as the servers used for testing.

4 TESTS AND RESULTS

The tests were done by running different number of jobs in parallel with several virtual machine configurations. High parallelism increases randomness in the finishing times. The job in a job set that finishes last determines the runtime of the whole set. This affects the energy-efficiency of the test as the trailing part is run inefficiently. Without this trailing effect the runs with more load would be better, which means that tests with higher parallelism would get better results.

4.1 Synthetic tests

As a synthetic test we had Lapack benchmark that is enhanced version of Linpack and is more suitable for a multicore environment. In our test we use Lapack to simulate a pure CPU-load. Test were run on single CPU server with powerful processor. Same tests were also run on energy-efficient processor, which resulted in similar curves. Main difference was in the minimum of energy-consumption and the maximum throughput. Energy consumption on energy-efficient processor was about 16% lower and throughput was about 31% lower.

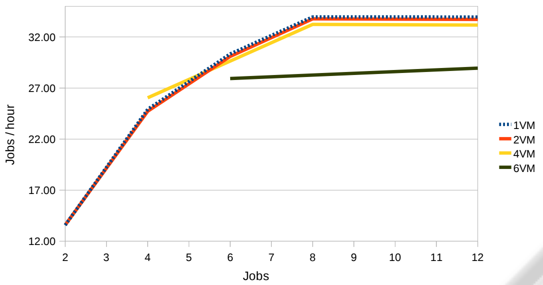

Figure 2: Lapack job throughput with different number of virtual machines and job parallelism.

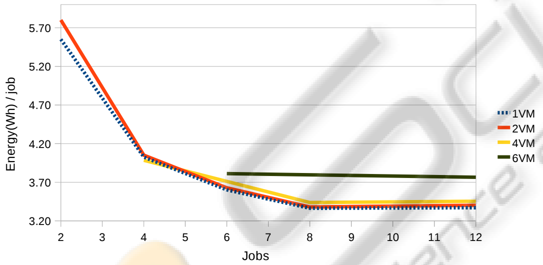

Figure 3: Lapack job energy consumption with different number of virtual machine and job parallelism.

Results from the synthetic tests indicate that the pure CPU-load is not affected that much by different virtual machine configurations. In Figure 2 we have the throughput results of the looped Lapack tests, which show that with one, two and four virtual machines there is no clear difference. When the number of virtual machines grow over the physical core count the throughput drops. Figure 3 shows the same behaviour in energy consumption per job. In both figures the number of jobs per virtual machine is the total number of jobs divided by the number of virtual machines.

4.2 Physics Tests

First we tested how the number of virtual machines affect throughput and energy efficiency of the virtualization host by running several virtual machines with one job each. In our earlier studies we have noticed that the commonly used one job per CPU core does not give the best performance or energy efficiency (Niemi et al., 2009b; Niemi et al., 2009a). Here we tested how it applies to virtualized environments.

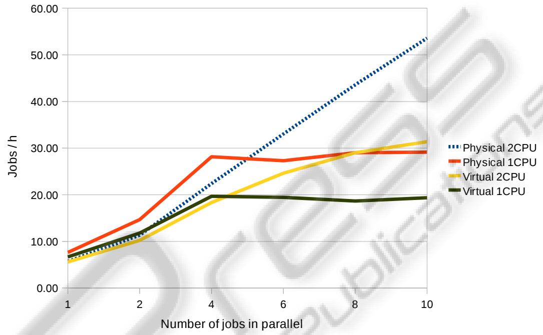

Figure 4: Throughput with different number of virtual machines running one job each.

In figures 4 and 5 we have results of running one job per virtual machine with increasing number of virtual machines. Results show that increasing the number of parallel jobs has more effect on the virtualized system, which saturates much earlier than a system without virtualization. The line of the two CPU server is cut before 12 jobs, but the two CPU server experiences same performance roof as the single CPU server with 4 cores.

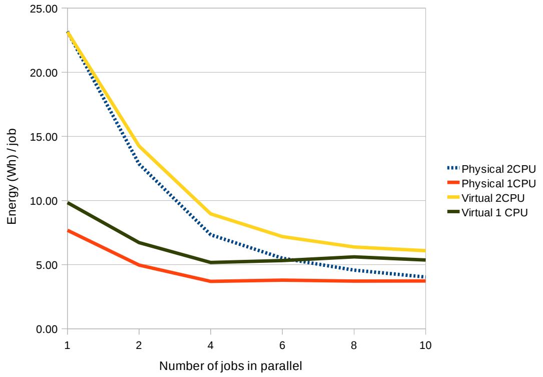

Figure 5: Energy consumption per job with different number of virtual machines running one job each.

Next we tested the servers with varying number of virtual machines and rising load. These tests also show how the energy-efficiency depend on how virtual machines share the load. The energy-efficiency declines when the number of virtual machines is increased and improves when the load in virtual machines is increased. In all figures the number of jobs in each virtual machine is the total number of jobs divided by the amount of virtual machines.

Figures 6, 7 and 8 illustrate results of the same physics test with three different processors and show that the effect of load division is the same with every hardware. Results show that both in synthetic and physics tests the throughput and energy-efficiency improves when the system load is increased.

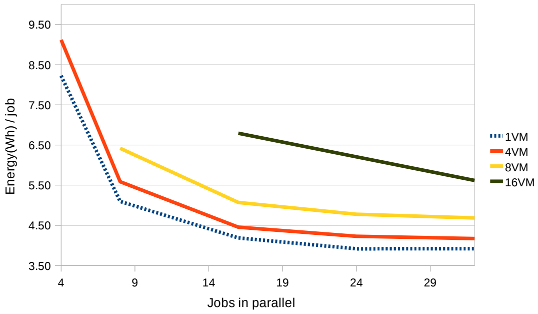

Figure 6: Energy usage per job with two CPU server.

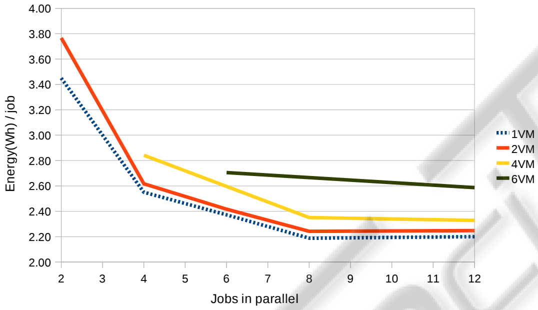

Figure 7: Energy usage per job with different VM parallelism on single CPU server with powerful processor.

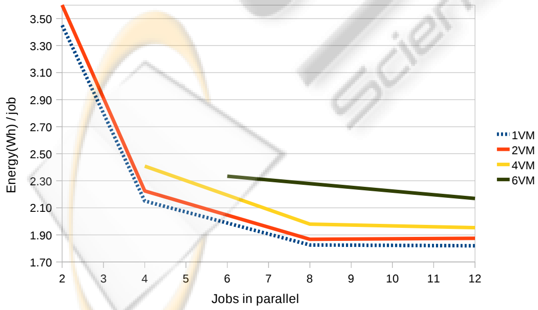

Figure 8: Energy usage per job with different VM parallelism on single CPU server with energy-efficient processor.

Results in Figures 7 and 8 show the results of test made with the two different single processor servers. These results are in line with the ones from synthetic tests. In the physics test the energy-efficient processor was 17% more energy-efficient and the throughput of the powerful processor 40% better.

In the physics test the number of virtual machines have a bigger effect on the virtualiation overhead as in the synthetic tests. Figure 9 shows how the overhead of virtualization increases exponentially when the number of virtual machines is increased. This overhead curve comes from running 12 jobs with different virtual machine sets. Result from running the 12 jobs in one virtual machine is compared to the results of running the 12 jobs in two, three and four virtual machines. In the case of four virtual machines one virtual machine runs three parallel jobs.

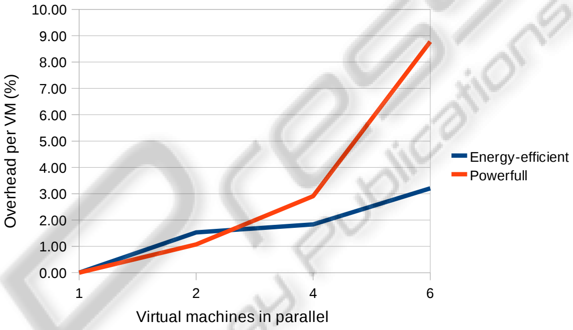

Figure 9: Energy overhead of virtualization per virtual machine with 12 jobs.

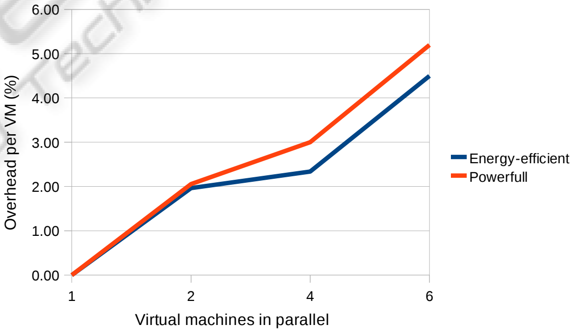

Figure 10: Duration overhead of virtualization per virtual machine with 12 jobs.

Figure 10 shows how the virtualization overhead increases the job duration. As did energy consumption, the duration increases exponentially in function of virtual machines. The results of the powerful processor show this better. The increase in overhead is not only caused by the processor core count as is occurs with two and four virtual machines.

5 CONCLUSIONS

The overhead of virtualization is well studied and there is many publications of it. Virtualization performance has improved from its early times and now it is in many cases very close to hardware level. There are still many technological challenges that need to be studied to improve virtualization performance, but even now it provides a useful platform for multitude of applications and is irreplaceable tool for energy efficiency in data centers.

We studied the energy-efficiency of virtualization and how virtual machine parallelism and load variation affects it. Our research indicates that the best energy efficiency and system throughput is achieved when the load is shared by small number of virtual machines. This depends much on the applications that are run in the virtual machines. Pure CPU-load in larger VM groups does not seem to impose as much overhead as the more complex physics analysis job. The physical core count also seem to pose a limit for the virtual machine pool.

REFERENCES

Anderson, E., Bai, Z., Dongarra, J., Greenbaum, A., McKenney, A., Du Croz, J., Hammerling, S., Demmel, J., Bischof, C., and Sorensen, D. (1990). Lapack: a portable linear algebra library for high-performance computers. In *Proceedings of the 1990 ACM/IEEE conference on Supercomputing*, Supercomputing '90, pages 2–11, Los Alamitos, CA, USA. IEEE Computer Society Press.

Antcheva, I. and et al. (2009). Root a c++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, 180(12):2499 – 2512.

Buyya, R., Yeo, C. S., and Venugopal, S. (2008). Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on*, pages 5 –13.

Chaudhary, V., Cha, M., Walters, J., Guercio, S., and Gallo, S. (2008). A comparison of virtualization technologies for hpc. In *Advanced Information Networking and Applications, 2008. AINA 2008. 22nd International Conference on*, pages 861 –868.

Fabozzi, F., Jones, C., Hegner, B., and Lista, L. (2008). Physics analysis tools for the cms experiment at lhc. *Nuclear Science, IEEE Transactions on*, 55:3539–3543.

Fenn, M., Murphy, M. A., and Goasguen, S. (2009). A study of a kvm-based cluster for grid computing. In *Proceedings of the 47th Annual Southeast Regional Conference*, ACM-SE 47, pages 34:1–34:6, New York, NY, USA. ACM.

Kivity, A., Lublin, U., and Liguori, A. (2007). kvm : the linux virtual machine monitor. In *Proceedings of the Linux Symposium*, pages 225–230.

Kommeri, J., Niemi, T., and Helin, O. (2012). Energy efficiency of server virtualization. In *Proc. Energy 2012*.

Niemi, T., Kommeri, J., and Ari-Pekka, H. (2009a). Energy-efficient scheduling of grid computing clusters. In *Proceedings of the 17th Annual International Conference on Advanced Computing and Communications (ADCOM 2009), Bengaluru, India*.

Niemi, T., Kommeri, J., Happonen, K., Klem, J., and Hameri, A.-P. (2009b). Improving energy-efficiency of grid computing clusters. In *Advances in Grid and Pervasive Computing, 4th International Conference, GPC 2009, Geneva, Switzerland*, pages 110–118.

Nussbaum, L., Anhalt, F., Mornard, O., and Gelas, J.-P. (2009). Linux-based virtualization for hpc clusters. *Network*, pages 221–234.

Padala, P., Zhu, X., Wang, Z., Singhal, S., and Shin, K., G. (2007). Performance evaluation of virtualization technologies for server consolidation. *Work*, (HPL-2007-59):15.

Regola, N. and Ducom, J.-C. (2010). Recommendations for virtualization technologies in high performance computing. In *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pages 409–416.

Schäppi, B., Bellosa, F., Przywara, B., Bogner, T., Weeren, S., and Anglade, A. (2007). Energy efficient servers in europe. Technical Report October, Austrian Energy Agency.

STAR, E. (2007). Report to congress on server and data center energy efficiency. Technical report, U.S. Environmental Protection Agency ENERGY STAR Program.

Verma, A., Ahuja, P., and Neogi, A. (2008). Power-aware dynamic placement of hpc applications. In *Proceedings of the 22nd annual international conference on Supercomputing*, ICS '08, pages 175–184, New York, NY, USA. ACM.

Xu, M., Hu, Z., Long, W., and Liu, W. (2004). Service virtualization: Infrastructure and applications. In *The grid: blueprint for a new computing infrastructure*, chapter 14. Wiley.