Experience with PROOF-Lite in ATLAS data analysis

S.Y. Panitkin[1], C. Hollowell[1], H. Ma[1], and S. Ye[1] for the ATLAS Collaboration

[1]Brookhaven National Laboratory, Upton, NY 11973, USA

Abstract. We discuss our experience with PROOF-Lite in a context of ATLAS Collaboration physics analysis of data obtained during the LHC physics run of 2009-2010. In particular we discuss PROOF-Lite performance in virtual and physical machines, its scalability on different types of multi-core processors and effects of multithreading. We will also describe PROOF-Lite performance with Solid State Drives (SSDs).

1. Introduction

PROOF-Lite is an implementation of the Parallel ROOT Facility (PROOF) [1] intended for an individual user running on a single machine. It is distributed with the Root [2] framework and requires zero configuration to run on multiple CPU cores. Like PROOF it allows one to increase performance of physics analysis by exploiting event level parallelism on multicore, multithreaded CPUs. It provides a straightforward scalability path for physics analysis, since software capable of running in PROOF-Lite on a single machine can run on a large scale PROOF farm without much modification.

2. PROOF-Lite in Virtual Machines

An interesting use case emerged when physicist from the ATLAS Collaboration [3] started to analyze LHC data in 2010 at the ATLAS Computing Facility (ACF) at Brookhaven National Laboratory (BNL). Initially they used PROOF-Lite based analysis on interactive nodes at ACF. These nodes are available to ACF users for code development and interactive analysis. The interactive nodes are actually Xen [4] virtual machines (VMs) in two different configurations. One set of 8 machines - named acas000[1-8] are 2 core VMs mapped to 2 physical cores. The remaining 4 machines, named acas00[09-12], are VMs with 4 logical cores mapped to 4 physical cores. All VMs run on similar underlying hardware, hypervisor and OS.

Questions about analysis performance on different interactive nodes arose during data analysis. This prompted a study of PROOF-Lite performance on modern software and hardware platforms with a goal to determine the optimal choice of hardware and software for ATLAS physics analysis at BNL. PROOF-Lite was tested on different virtual and physical machines. In order to study the dependence of analysis rate on the number of cores in a virtual machine, VMs were configured with 2, 4 and 8 cores.

Each physical machine hosting VMs has dual quad core Intel Xeon CPUs (E5335) running at 2 GHz and a 1Gbps network interface. All VMs were configured with 8GB RAM. At the time of
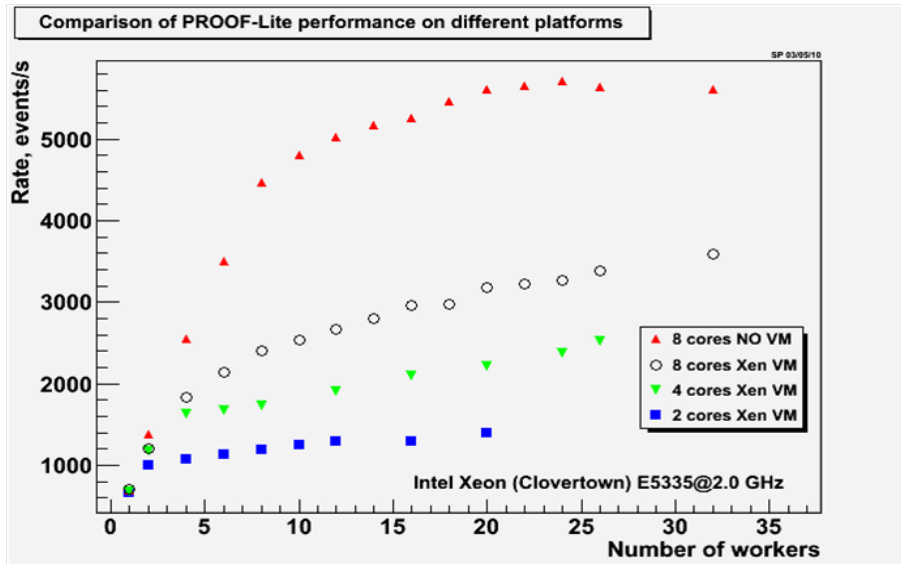
Figure 1. Analysis rate as a function of the number of PROOF-Lite workers for VMs with different number of cores. See description in the text

our study VMs ran x86_64 kernel 2.6.18-164.2.1.el5xen.

Typically, remaining physical cores on the same node are used by another VM which is configured to be a part of a Condor batch queue. For our tests the neighboring batch VM was disabled to reduce interference with measurements.

The test jobs were run on a subset of 900 GeV data taken by the ATLAS Collaboration at LHC during the 2009 and 2010 runs (ATLAS dataset identifier- DATA900_MinBias_esd_r998_V17). This subset contained 208 files of ATLAS Event Summary Data (ESD) in D3PD format. It had 1103478 events, with a total size of 19 GB, and average event size of 17.4 kB/event. Files were located on one of the ACFs high-performance BlueArc Mercury 100 [5] NFS servers. Test jobs were examples of ATLAS min-bias analysis code.

Figure 1 shows results of the tests for VMs with different number of cores. We found, perhaps not surprisingly, that VMs with a large number of available cores provide better performance. For comparison we also showed in Figure 1 analysis performance for an equivalently configured physical machine with 8 cores. It is clear from the Figure that our analysis jobs ran significantly faster in a non-virtualized environment.

We would like to note that results of tests in virtual environment will depend on many factors like workload type, hypervisor flavor (Xen, VMWare, KVM), tuning of the guest VM, etc. We did not attempt to study influence of all these factors on analysis performance. Our tests were done in a production environment typical for ACF at BNL.

3. PROOF-Lite on Different CPUs

During 2010 ACF was undergoing a hardware upgrade so we had a chance to compare PROOF-Lite performance on different CPUs. Comparison between X5550 Nehalem and E5335 Cloverton CPUs was of particular interest. They represent subsequent CPU generations utilized at ACF. Of course the difference between the generations is not only in CPU but also in different chipset and memory architectures. In this document we will use CPU models as labels to denote different computing platforms. Machines that were involved in the comparison were configured with an identical amount of RAM - 24 GB, and equipped with a 1 Gbps network card. The
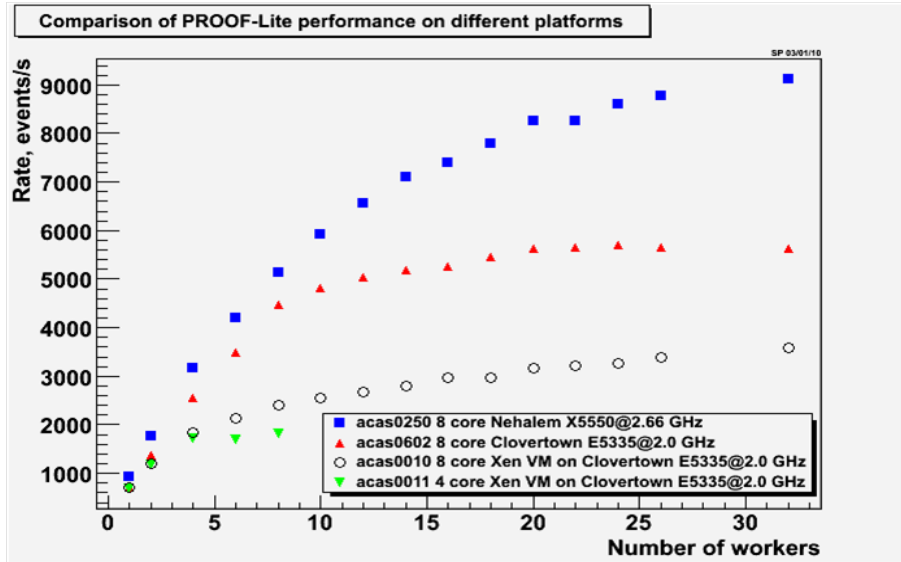
Figure 2. Analysis rate as a function of the number of PROOF-Lite workers for different CPUs. See description in the text

same dataset described in the previous section, located on a BlueArc NFS server, was used for these tests. We used the same code as mentioned in the previous section. Figure 2 shows a comparison between PROOF-Lite performance scaling on the two CPU platforms. For the sake of comparison we also show in Figure 2 PROOF-Lite performance for VMs running on Cloverton based hardware. Labels acas[*] in Figure 2 refer to test machine names as assigned by ACF. Not surprisingly the X5550 Nehalem platform shows better performance in our tests. In addition to having a newer CPU architecture, it is also clocked higher. Is the performance difference mostly determined by the CPU frequency scaling? If frequency scaling was the only determining factor then one would expect performance ratio to be constant and close to the ratio of CPU clock frequency ($\sim$1.3). Figure 3 shows the ratio of analysis rates for the two platforms. It is clear from the Figure that CPU frequency scaling does not tell the complete story. The Nehalem architecture (X5550) offers better performance for a large number of workers, but the Cloverton architecture is quite competitive for a smaller number of workers.

3.1. Effect of Hyperthreading

The X5550 CPU provides hardware support for hyperthreading. We looked at what effect hyperthreading has on our workload. For this test we used the same code and data as mentioned in previous sections. Figure 4 shows the analysis rate as a function of the number of PROOF-Lite workers for the X5550 CPU with hyperthreading turned on and off. Hyperthreading does not seem to make a big impact on performance in our case. With a small number of PROOF-Lite workers it seems to reduce the analysis rate a little, but when the number of workers exceeds the number of physical cores in the system, then hyperthreading gives a small boost in performance compared to a non-hyperthreaded analysis case. In both cases the effect is small, less than 5%.

4. PROOF-Lite with SSD

We tested PROOF-Lite performance with data on Solid State Drives (SSD). An earlier comparison between regular hard drives and SSDs in PROOF based analysis was published
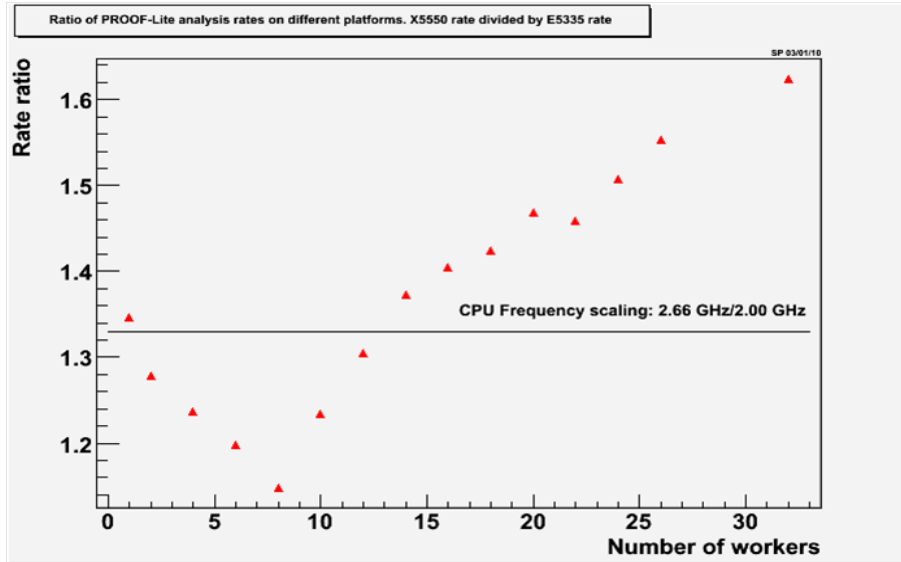
Figure 3. Ratio of analysis rates for X5550 and E5535 CPUs. See description in the text
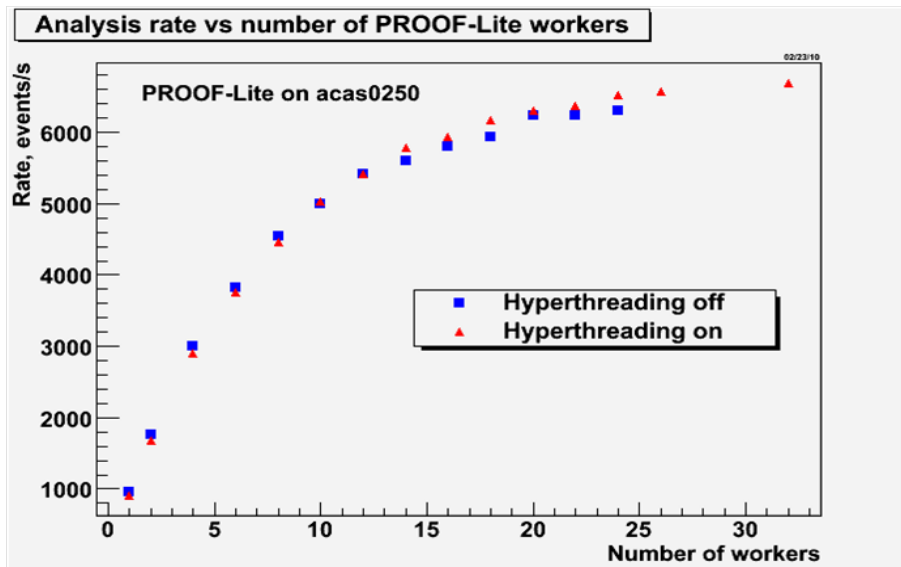

Figure 4. Analysis rate as a function of the number of PROOF-Lite workers for the X5550 CPU with hyperthreading turned on and off. See the description in the text

in the CHEP09 proceedings [6]. It was noted there that SSD technology provides an excellent platform for data analysis and offers significant performance advantages over conventional hard drives in an I/O dominated physics analysis environment.

For PROOF-Lite tests we have used Mtron model MSP-SATA7035064 drives, the same drives utilized in [6]. This time we compared PROOF-Lite performance for the case where data is on an SSD software RAID against the case where data was placed on a BlueArc Mercury NFS server. We used the same dataset and analysis code as discussed in previous sections. For SSD tests the dataset was copied to SSDs and, before each test run, the test machine was rebooted to ensure
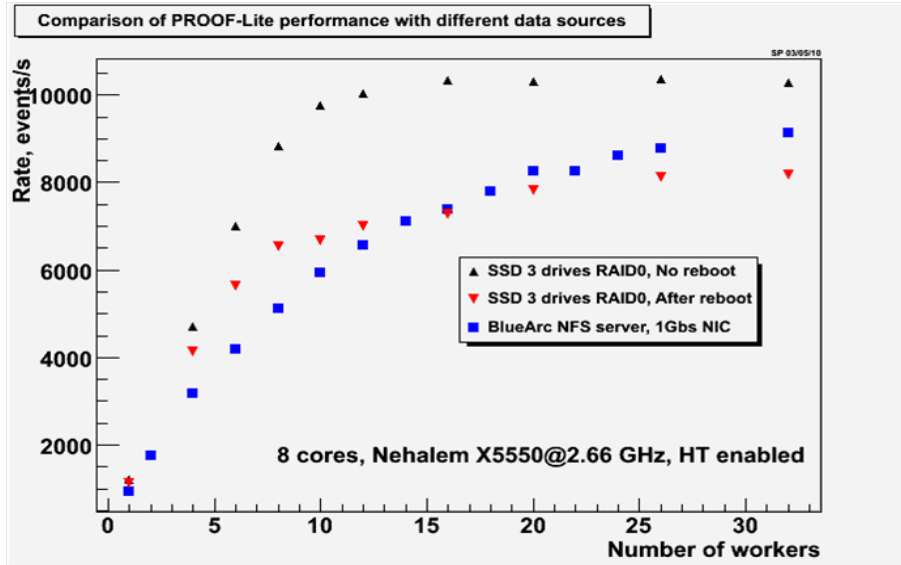
Figure 5. Analysis rates as a function of the number of PROOF-Lite workers for data on an SSD RAID array and a BlueArc Mercury NFS server. See description in the text

that effects of data caching in RAM did not affect consecutive tests results. Figure 5 shows analysis rates as a function of the number of PROOF-Lite workers for data on a 3-drive SSD RAID0 and a BlueArc Mercury NFS server. For comparison we also show, in black triangles, PROOF-Lite scaling for the case when the analysis dataset was memory resident. For this case it is clear that for a given work load our tests system starts to run out of CPU cycles at about 10-12 workers and becomes fully CPU limited at about 16 workers. The three drive SSD array showed better performance, compared to BlueArc, for a small number of workers (less than 16), but BlueArc NFS server performance was better at large number of Proof-Lite workers. This may be explained by a combination of factors. With a number of PROOF-Lite workers larger than 8 the SSD analysis rate increase diminishes since at this point the system seems to become largely I/O limited. Remaining CPU headroom allows gradual though slower performance increase. For BlueArc case higher latencies of data access over the network, in combination with large available bandwidth [5], allows for slower, more gradual, increase in analysis rate sustainable up to 32 workers.

5. Summary

In our experience PROOF-Lite is an excellent tool for physics analysis. It allows one to easily harness the processing power of modern multicore CPUs. It is being actively used in ATLAS data analysis at BNL. We performed a series of tests in order to determine the optimal software and hardware setup for PROOF-Lite based ATLAS data analysis in the ACF environment. We found that for our analysis cases PROOF-Lite suffers a significant performance penalty when running in a Xen virtual machine. Our PROOF-Lite based analysis did not benefit much from CPU hyperthreading on the Nehalem platform. Solid State Drives provide very good performance for physics analysis. We also note that the BlueArc Mercury NFS server showed good performance as a data source for our PROOF-Lite based analysis.

References
[1] Ballintijn M *et al.* 2006 *Nucl. Instrum. Meth.* **A559** 13–16

[2] Brun R, Rademakers F, Canal P and Goto M 2003 (*Preprint* cs/0306078)

[3] Aad G *et al.* (ATLAS) 2008 *JINST* **3** S08003

[4] Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauer R, Pratt I and Warfield A 2003 *SIGOPS Oper. Syst. Rev.* **37**(5) 164–177 ISSN 0163-5980 URL http://doi.acm.org/10.1145/1165389.945462

[5] 2010 Bluearc mercury network storage stystem URL http://www.bluearc.com/bluearc-resources/downloads/data-sheets/BlueArc-SS-Mercury.pdf

[6] Panitkin S *et al.* 2009 Study of solid state drives performance in proof distributed analysis system *Proccedings of International Conference on Computing in High Energy Physics (CHEP 09)*