An Overview of the Options Available for Practical Activities in Distributed Computing

Camille Coti†‡ camille.coti@univ-paris13.fr
Jean-Vincent Loddo†‡ jean-vincent.loddo@univ-paris13.fr
Emmanuel Viennet†♮ emmanuel.viennet@univ-paris13.fr
† : *IUT de Villetaneuse, Université Paris 13, Sorbonne Paris Cité*
‡ : *LIPN, CNRS UMR 7030, Institut Galilée*
♮ : *L2TI, Institut Galilée*

Practical activities in computer science often require a computer. For this purpose, lab rooms feature computers, roughly one for each student in the room. High-performance, parallel and distributed computing is particular in a sense that it requires *several* computers for each student. In this paper, we will examine several solutions that can be used for training in high-performance computing and distributed systems to provide students with realistic conditions for practical activities.

a) Comparison criteria: First, we will briefly analyze what is needed for practical activities in order to reproduce realistic conditions for the students. This part helps us comparing the appropriateness of the various solutions presented in the rest of the paper. Students can need to work on distributed resources in two main categories of contexts: administration of clusters and data centers, and use and programming of distributed resources. These two context focus on different aspects of distributed resoures, and therefore the features provided by the lab platform are different regarding the goal of the practical activities.

We will enumerate various solutions that we have used with our students at both undergraduate and graduate level, their feasibility (from both practical and administrative points of view) and how realistic the result is in terms of usage and feedback.

b) Categories of use-cases for practical activities: We identified the various themes for practical activities and the associated needs in two categories. It can be about *distributed programming*, *i.e.* programming applications that run on several machines. It includes parallel programming (such as MPI programming), cloud computing (Hadoop and other MapReduce implementations, NoSQL databases...), client-server programming, or any kind of distributed applications. On the other hand, it can be about *network and system administration*, such as the administration of network services (DHCP, DNS, distributed file systems, email service, access directories...) or datacenter administration procedures.

c) Solutions compared: The second part of this paper will compare the solutions themselves with respect to what can be required by the practical activities themselves. We will compare *running multiple processes on a single machine*, which is not very realistic but does not require any hardware nor software adaptation; *virtualization*, such as Marionnet, which is very realistc but has some performance issues; *using the machines of the lab room*, which is a good solution but the performance (and therefore how realistic it is) depend on how many students at a time are using the machines, and moreover in the general case the students do not have root access on the machines; *using a cluster*, which is realistic for HPC classes but not other kinds of classes, and the habits of cluster usage include using a batch scheduler in asynchronous mode, therefore making the executions asynchronous, even with queues with higher priorities; *using a virtual, teaching cluster* on dedicated resources is a better solution than the previous one, but the performance depends on the virtualization solution used and requires some dedicated hardware, which has a cost; *using a cluster of single board computers* for each student or group of students, such as Raspberry Pi computers, which is not very expensive but still has a cost, and is not very performant in terms of computation capabilities and network interconnection; *renting resources on a cloud*, such as EC2, which is realistic, has good performance, but can be an administrative nightmare in some departments and can be expensive.

d) Conclusion: To conclude this paper, we summarize what the most appropriate solutions are for each group of practical activities, and we give our personal feedback with the solutions we have adopted for the various classes we are teaching on this topic.

This paper was accepted at ITHET 2016. However, due to the serious events that happened in Turkey just before ITHET 2016, the French government has put Turkey on a list of countries for which government employees need to get an official clearance from a Defense and Security Official before they apply for a mission order in the country. The short timeframe combined with the uncertainty of the political situation in Turkey made it impossible for us to apply in time for this clearance and we could not make it to ITHET. Regarding the situation of last year's conference, can we transfer our paper from last year into ITHET 2017?

REFERENCES

[1] J.-V. Loddo, L. Saiu: *Marionnet: a virtual network laboratory and simulation tool*, 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SimulationWorks'2008), Marseille, France, 2008.

[2] J.-V. Loddo, L. Saiu, *Status report: marionnet or "how to implement a virtual network laboratory in six months and be happy"*, ACM SIGPLAN Workshop on ML (ML'2007), Freiburg (Germany), 2007.

[3] J. Dike: *User mode linux (Vol. 2)*, Englewood Cliffs: Prentice Hall, ISBN 0-13-186505-6, 2006.

[4] R. Davoli: *VDE: Virtual Distributed Ethernet*, First International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, (TRIDENTCOM 2005), pp. 213-220, IEEE, 2005.

[5] C. Braun: *Mobile clusters of single board computers: an option for providing resources to student projects and researchers*, pp. 1-20, SpringerPlus, 2016, vol. 5, no 1.

[6] C. Coti, J.-V. Loddo, E. Viennet: *Practical activities in network courses for MOOCs, SPOCs and eLearning with Marionnet*, Proceedings of the 14th International Conference on Information Technology Based Higher Education and Training (ITHET15), June 2015.