

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

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

I. INTRODUCTION

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.

We enumerate different 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.

II. CATEGORIES OF USE-CASES FOR PRACTICAL ACTIVITIES

Labs involving a distributed system can be divided into two categories: those about *using* a distributed system, and those about *making* a distributed system. Their expectations and how they use the distributed system determines what they can use to have a realistic feeling, *i.e.* how the lab platform can make the students feel like they are using something close to a real system.

A. Using a distributed system

When students are using a distributed system, they have access to a set of resources that communicate together. For instance, distributed, parallel programming (such as MPI programming) and cloud computing (Hadoop and other MapReduce implementations, NoSQL databases..) require using a set of computation resources and a middleware that orchestrates them.

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?

B. Distributed system programming

Distributed system programming includes programming applications that execute on several machines: client-server applications, network programming, implementation of distributed algorithms... Some of them might require that only one instance of each process is executed on a given machine: for instance, if all the processes are using the same network port, if they need to use a central program such as RPC's portmapper...

C. Distributed system administration

III. COMPARISON CRITERIA

A. Realistic performance

B. Realistic procedures and system

C. Feedback latency

The last criterion we are considering is how quickly the students get feedback for their work.

IV. SOLUTIONS COMPARED

A. Running multiple processes on a single machine

In lab rooms, students work on a computer. The simplest idea is to run multiple processes on a single machine. Only one machine is needed, and all the operations are performed on it.

Even though it is the simplest implementation, this solution is not very realistic. If the number of processes executed by the distributed system is larger than the number of cores available on the machine, the system will be (partially) sequentialized. Moreover, some systems might need to access resources for which only one instance is available on the machine, *e.g.* ports or IO devices, will cause conflicts between the processes.

B. Virtualization

TODO

for instance, Marionnet - Realistic - Performance issues (citation Ben HPDC 2006)

C. Using the cluster

- Most realistic for HPC, not the other ones - Queuing issue
- Asynchronous submissions: not sure when the result will arrive, even with queues w/ higher priority - Non-interactive

D. Virtual, teaching cluster

Dedicated resources - better than the previous solution - performance issue, depends on the virtualization solution - cost!

E. Renting resources on a cloud (like EC2)

- realistic - performance - cost - administrative mayhem

F. Using the computers of the lab room

- performance vary with usage - not always root

G. Using a cluster of single board computers

V. CONCLUSION

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 (Simulation-Works'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.
- [7] Benjamin Quetier, Thomas Herault, Vincent Neri, Franck Cappello: *Virtual Parallel Machines Through Virtualization: Impact on MPI Executions*. Proceedings of EuroPVM/MPI 2007: 381-383, September 2007.
- [8] Lamia Youseff, Rich Wolski, Brent Gorda, Chandra Krintz: *Evaluating the Performance Impact of Xen on MPI and Process Execution For HPC Systems*. In VTDC@ SC (p. 1), November 2006.