Condor and Philosophy of Flexibility: A Case Study of Two Scientific Collaborations Using High-Throughput Computing

Sara Kraemer and Chris Thorn

Wisconsin Center for Education Research
University of Wisconsin-Madison
sbkraeme@wisc.edu, cathorn@wisc.edu

Abstract. In this paper, we summarize two scientific collaborations using Condor at the University of Wisconsin-Madison, a distributed High-Throughput Computing (HTC) resource that supports a range of scientific collaborations, including cross-disciplinary, virtual and co-located teams, among others. The Condor project embodies a 5-component philosophy of flexibility that addresses the barriers to unpredictable distributed operating environments. We interviewed 2 scientific collaborations using Condor as the primary source of HTC resources to describe how the philosophy of flexibility works within each of the collaborations. We have adopted a sociotechnical approach to investigating these teams and outlined some preliminary characteristics of HTC teams using Condor.

1 Introduction

An increasing range of scientific research is addressing problems that can best be addressed using computationally-intensive analytical tools. High-throughput computing (HTC) addresses this need by supplying large amounts of computing power over distributed networks. We define HTC as an environment that can deliver large amounts of processing capacity over long periods of time. In addition to computational cycles delivered, there is a second, critical measure of system

quality: HTC systems are designed to be extremely fault-tolerant and require minimum human intervention (Thain et al., 2005). By design, these technologies enable and support distributed teams. These and other characteristics channel the interactions and forms of collaboration that emerge when users from various scientific domains use HTC resources to work on computational problems.

Scientists in many disciplines have begun revolutionizing their fields by using HTC resources in technology-mediated, distributed-work environments. Some of these trends have included the extension of complex simulation and modeling from classic approaches to scientific research (i.e., theoretical/analytical, experimental/observational) (Atkins et al., 2003). In addition to these new trends in computational science, one of the central HTC considerations is sociotechnical implications of shared computational resources, ownership, and cooperation (Thain et al., 2003).

Some HTC systems, such as the Condor$^©$ project at University of Wisconsin-Madison, have unique characteristics to address these implications, such as high degrees of resource flexibility, end-user control, open-ended planning, and distributed resource management (Thain et al., 2005). The Condor project embodies a *philosophy of flexibility*, which has allowed HTC systems to flourish in highly unpredictable distributed operating environments (Thain et al., 2005).

In this paper, we will explore the sociotechnical characteristics of Condor's philosophy of flexibility with 2 case studies of scientific collaborations at the Grid Lab of Wisconsin (GLOW). GLOW is a cluster of cooperative computing resources using Condor at the University of Wisconsin-Madison (GLOW, 2009). We will first describe the overall GLOW project and Condor's philosophy of flexibility followed by a case study analysis of two GLOW teams.

2 GLOW, Condor, and Philosophy of Flexibility

GLOW is a distributed scientific computing resource at the University of Wisconsin-Madison that combines and enhances autonomous sites of computing resources. GLOW is an interdisciplinary effort that spans 10 scientific domains: Biostatistics and Medical Informatics, Chemical and Biological Engineering, Chemistry, Computer Sciences, Engineering Physics, Genomics, Genetics, Materials Science and Engineering, Medical Physics, Physics, and Astrophysics. The laboratory consists of 8 physical sites and individually provides the necessary hardware, software, and support infrastructure for the development and experimental evaluation of HTC applications. Each of the sites focuses on addressing local computational needs and maintains full control over local resources while sharing unused computing power and storage space across site boundaries according to a group defined policy. The goal of the *laboratory* is to bring together domain and computer scientists to make HTC computing an effective tool for scientific research by harnessing and sharing the power of

commodity resources. GLOW members collaborate in the development, implementation, testing, and deployment of grid-enabled capabilities while cultivating interdisciplinary science.

The underlying computational cyberinfrastructure and resource management for GLOW is supplied by the Condor project. Condor has characteristics that differentiate it from other HTC resources. For example, because Condor runs on many computing platforms and operating systems and can execute any software that does not require user interaction. Condor also offers a wide range of tools that are readily available for users—from commercial research software to scripting engines and compilers. In addition, the available scientific tools allow individual scientists or teams to engage with the Condor HTC environment using tools familiar to them. Further, enabling of existing tools in an HTC setting provides critical social and technological gateways for new adopters of HTC. Access to the HTC environment also exposes new adopters to tools and methods used by others to address similar computational and/or analytical problems. In this way, scientists' skills and knowledge are affected by the capabilities and characteristics of HTC technologies and tools.

The Condor project embodies a philosophy of flexibility; this philosophy that has served the allowed the design to flourish in a highly unpredictable distributed operating environment (Thain et al., 2003). International distributed systems are heterogeneous in numerous ways: they are composed of many types and brands of hardware; they run various operating systems and applications; they are connected by unreliable networks; they change configuration constantly as old components become obsolete and new components are become online, and they have many owners with local policies and requirements that control their participation in the community. Condor has adopted a 5-component flexibility philosophy to address these barriers:

(1) *Let communities grow naturally.* Given tools of sufficient power, people will organize the computing structures they need. However, human relationships are complex, and people invest their time and resources to varying degrees and relationships and requirements change over time. Therefore, Condor design permits but does not require cooperation.

(2) *Leave the owner in control, whatever the cost.* To attract the maximum number of participants in a community, the barriers to participation must be low. Users will not donate their property to the "common good" unless they maintain some control over how it is used. Therefore, owners of computing resources are given the tools to set policies and retract resources for private use.

(3) *Plan without being picky.* Plan for slack resources as well as resources that are slow, misconfigured, disconnected, or broken. The designers of Condor

spend more time and resources contemplating the consequences of failure than the potential benefits of success.

(4) *Lend and borrow*. The Condor project has developed a large body of expertise in distributed resource management and aims to give the research community the benefits of their expertise while accepting and integrating knowledge and software from other sources. They have also instituted a mechanism for collective problem-sharing and solving among its users.

(5) *Understand previous research*. The Condor project continually updates its organizational knowledge with previous research to apply well-known fundamentals as well as cutting-edge techniques to emergent problems. The inclusion of current user innovations keeps the work focused on the edge of discovery rather than wasting effort remapping known territory.

As outlined in the philosophy of flexibility, the Condor approach is more than a complex set of computational resources. The Condor team maintains a close intellectual partnership with GLOW teams and works together on the challenges of HTC in the context of break-through science. Condor has advanced HTC technology via improvements in their software coupled with innovations in the computational approaches with the domain scientists. These interactions have made Condor privy to numerous sets of interdisciplinary virtual/co-located teams as well as numerous types of sociological and technological factors encountered in research settings. In this next section, we will highlight some of these factors in 2 GLOW teams: IceCube and the Laboratory for Molecular and Computational Genomics.

3 GLOW Teams: IceCube and The Laboratory for Molecular

We have conducted group interviews with 2 GLOW teams to assess how Condor's philosophy of flexibility is implemented in live research settings (see Appendix for interview guide). The first team is IceCube, a collaboration supporting a neutrino detector at the South Pole; the second team (lab) is the Laboratory for Molecular and Computational Genomics (LMCG) at the University of Wisconsin. The IceCube group interview consisted of the P.I. and computational resource scientist who manages the HTC for IceCube group at UW-Madison. The LMCG group interview consisted of the P.I., 2 research scientists, and 1 post-doc.

UW-Madison is the lead institution for the construction and operation of IceCube, as well as the largest group of faculty, scientists, post-docs, and students in the international IceCube collaboration of over 250 people in 35 institutions. IceCube, a telescope under construction at the South Pole, will search for

neutrinos from the most violent astrophysical sources: events like exploding stars, gamma ray bursts, and cataclysmic phenomena involving black holes and neutron stars. The IceCube telescope is a powerful tool to search for dark matter, and could reveal the new physical processes associated with the enigmatic origin of the highest energy particles in nature. IceCube encompasses a cubic kilometer of ice and uses a novel astronomical messenger called a neutrino to probe the universe. (IceCube Neutrino Observatory, 2009).

GLOW's role has touched all aspects of commissioning and operation of the detector, especially in the challenging production of real-time detector simulations. The scientific analyses and simulations of for UW-Madison scientists rely completely on GLOW resources; these analyses and simulations will also completed for other collaborators on the IceCube team once the operational phase of the detector is fully employed.

GLOW has assisted in the timeliness of delivering the results of simulations and experiments to IceCube team members, which would yield a competitive advantage. Further, since the IceCube collaboration is about ¼ the size of typical particle physics collaboration they are able to more easily communicate, share, and work on findings which fosters an overall culture of cohesiveness (compared to typical particle physics collaborations). Further, the on-time delivery of data results has also increased the level of the group's productivity; because many tasks and data results are interdependent to ongoing work in other scientists.

The use of Condor has also emerged independently in several IceCube collaborating institutions. Some of their European collaborators share access to distributed researchers with groups using grid technologies.

The LMCG investigates single molecule phenomena for the creation of new systems in the biological sciences (LMCG, 2009). The size of the LMCG is smaller, 12-13 people in this lab, and the entire lab is located at UW-Madison. Within this group, many disciplines are represented; including but not limited to chemistry, statistics, bioinformatics, engineering, and genetics. The LMCG team has also had an in-house computational resource "expert" facilitate the work between Condor and LMCG (although this position is currently vacant). They felt that this facilitator role was an immense resource for identifying LMCG's scientific needs and effectively translating them into computational requirements for Condor. This role also proactively anticipated any problems or potential concerns, and mitigated them on behalf of the LMCG to Condor. LMCG attributed the accomplishment of many large research goals to close collaboration between the LMCG facilitator and the Condor team of experts.

3.1   Philosophy of Flexibility in GLOW

Let communities grow naturally, leave the owner in control, plan without being picky

GLOW was formed as a collective effort of domain scientists to share computational resources at UW-Madison in collaboration with computer scientists at Condor. They have a shared interest in consuming a large amount of computational resources for their research problems and therefore found that sharing slack resources would benefit their projects in key ways, such as computational speed, efficiency, and minimizing complexity. Because of the open-ended nature of the Condor philosophy, GLOW participants each have individual input for decision-making of the collective.

For example, a critical organizational design piece of GLOW is the monthly meeting of GLOW teams. At this monthly meeting, decisions about resource allocation and usage are made among the group; attendees are typically those who organize computational resources for each physical site. This meeting was described as a time for "allocating opportunistic time". In addition to these monthly meetings, individual research sites contact and request resources and expertise directly to the Condor team. The IceCube and LMCG projects both reported that their computational resource needs have always been met, although they may experience differences among GLOW team members about how resources should be divided (this is very rare). IceCube noted because of this intra-group structure and organization of available slack resources, there is very little to no competition among the collective for slack resources.

Lend and borrow, understand previous research

Both GLOW teams relied heavily on Condor's expertise in distributed resource management to both deliver the computational power needed to execute their scientific algorithms, models, and simulations, but also contributed to creative approaches to complete their work in ways that allowed the scientists to more efficiently and collaboratively. In addition, the Condor team both supports the use of Condor generally as well as engages very deeply with local clients' and their projects, supporting them in the defining and redefining of their research problems, assisting with interdisciplinary collaboration, and providing technical assistance, which was echoed at both sites. The LMCG noted that over the years-long collaboration with Condor, they have found working with the Condor system has progressively become easier to use and they attributed this improvement to continuous feedback Condor elicits from user groups as well as current research from in the field of study.

Some examples of Condor's help have included assistance re-tooling complex algorithms to run on Condor. At the LMCG, the computational resource staff person would interact with Condor as well as the domain scientists (a mathematician in this example) to figure out a way to re-tool the algorithm.

Further, because of the availability of computational resources, the revised algorithm did not have to be written in a way that conserved computational resources. As a result, the time needed to complete the new algorithm was lessened considerably. This scientist noted that Condor provided "the gift of time" to work on other problems.

4 Future Research

This paper summarizes some of our preliminary work in investigating how Condor's philosophy of flexibility manifests in interdisciplinary collaborations. We are currently studying other GLOW teams and aim to further explore how these tenets contribute to or hinder the effectiveness of teams' collaboration and scientific output. We also are very interested in developing our understanding of virtual team performance using Condor and the similarities and differences of those various teams, from a sociotechnical perspective. An analysis of the social impacts of the technical configurations of Condor, may lead to deeper understanding of how HTC is used as an effective enabler of new scientific problem sets, solutions, and collaboration configurations. In addition, this investigation can also influence the design of Condor, and HTC technology in general, to meet emerging scientific problems and configurations.

Appendix: Interview Guide

1. Do you think that using Condor and interacting with Condor team members has contributed to breakthroughs in your science?
   a. How has it specifically contributed to breakthroughs? Examples needed.
2. Does Condor allow you to organize your computational resources, and by extension, team and work, in a way that best reflects your research needs and project tasks?
   a. If not, why? If so, how? Examples.
   b. Has using Condor changed the way you work with other scientists, both within and outside of the team? How?
3. How do you set Condor policies and settings within your project?
   a. Do you feel any aspect of your work is affected by sharing computational resources with other Condor users?
   b. Does sharing Condor use facilitate or hinder any aspect of your science? Does it vary at all across different scientific disciplines?
4. Have you encountered any problems with using Condor as a computation resource? If so, how?

a. What are the best aspects of using Condor for virtual teamwork and interdisciplinary scientific collaboration?

5. Has interacting with the Condor team affected the quality of scientific output and/or the collaboration among team members?

b. If so, how? If not, why not? Examples?

Acknowledgments

5   References

Thain, D., Tannenbaum, T. and Livny, M. (2005): Distributed computing in practice: The Condor experience. Concurrency and Computation: Practice and Experience, 17 (2-4). 323-356.

Atkins, D.E., Droegemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L., Messerschmitt, D.G., Messina, P., Ostriker, J.P. and Wright, M.H. (2003): Revolutionizing science and infrastructure through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on cyberinfrastructure: National Science Foundation.

Thain, D., Tannenbaum, T., Livny, M., Berman, F., Fox, G. and Hey, T. (2003): Condor and the Grid. in D. Hutchinson (Ed.), Grid Computing: Making the Global Infrastructure a Reality (pp. 299-335): Johh Wiley & Sons.

GLOW. (2009): http://www.cs.wisc.edu/condor/glow/.

IceCube Neutrino Observatory. (2009): http://www.icecube.wisc.edu/info/.

LMCG. (2009): http://www.lmcg.wisc.edu/.