Lessons from an Online Multidisciplinary Undergraduate Summer Research Program

Matthias K. Gobbert[1] and Jianwu Wang[2]

[1] Dept. of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA, gobbert@umbc.edu
[2] Dept. of Information Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA, jianwu@umbc.edu

Abstract. During 2018, 2019, and 2020, the UMBC CyberTraining initiative "Big Data + HPC + Atmospheric Sciences" created an online team-based training program for advanced graduate students and junior researchers that trained a total of 58 participants. The year 2020 included 6 undergraduate students. Based on this experience, the authors created the summer undergraduate research program Online Interdisciplinary Big Data Analytics in Science and Engineering that will conduct 8-week online team-based undergraduate research programs (bigdatareu.umbc.edu) in the summers 2021, 2022, and 2023. Given the context of many institutions still being online and potentially expanding their online instruction in the wake of the COVID-19 pandemic, we share our experiences how the successful lessons from CyberTraining transfer to a high-intensity full-time online summer undergraduate research program.

Keywords: Online Education, Big Data, High-Performance Computing, Multidisciplinary Education, Team-Based Learning

1 Introduction

Next to theory and experimentation, computation has become the third pillar [1] and data-driven science has become the fourth pillar of the scientific discovery process [2] for many disciplines and critical to their research advances, such as bioinformatics, physics, computational chemistry, and mechanical engineering. It demands requirements on a training explaining how data and computation related techniques can help scientific discovery. Yet such a "Data + Computing + X" training is often missing in current curriculum design.

In 2017, the U.S. National Science Foundation (NSF) published the solicitation "Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)" designed to address this national need. This program continues currently with solicitation number NSF 19-524. Four faculty from three departments across two academic colleges at UMBC joined in response and proposed the UMBC CyberTraining initiative to create the nationwide online team-based training program "Big Data + HPC + Atmospheric Sciences" (cybertraining.umbc.edu) for students in three disciplines (Computing, Mathematics, and

Physics) to foster multidisciplinary research and education using advanced cyberinfrastructure (CI) resources and techniques. Figure 1 illustrates graphically the connection between the disciplines.

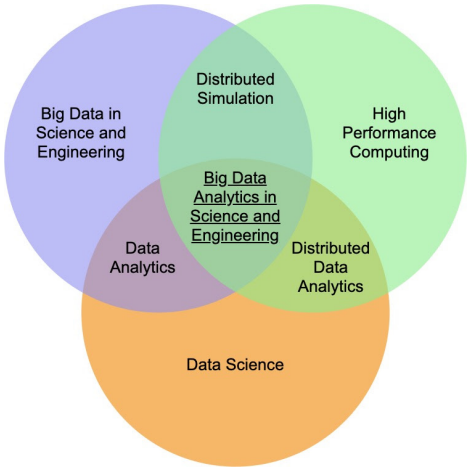

Fig. 1. Illustration of the connections between Big Data, HPC, and their applications in science and engineering.

2 The Big Data + HPC + Atmospheric Sciences Program

The CyberTraining program teaches participants how to apply knowledge and skills of high-performance computing (HPC) and big data to solve challenges in Atmospheric Sciences. We focused on the application area of atmospheric physics and within it radiative transfer in clouds and global climate modeling, since these topics are important, pose computational challenges, and offer opportunities for big data techniques to demonstrate their impacts. The NSF funded our proposal in the inaugural year 2017 (OAC–1730250) for training programs conducted in 2018, 2019, and 2020 [3].

This program trained 58 participants and we reported on our experiences in conducting such training online and team-based with participants ranging from undergraduates (NSF-funded through an REU Supplement in Year 3), graduate students, post-docs/non-TT faculty, and TT (tenure-track) junior faculty [4]. We specifically describe how to practically create the necessary training material, chiefly the tapings of lectures for later asynchronous online delivery of contents and homework, during Year 1, and how to accomplish this in an institutionally supportive environment, but without the type of resources an institution with an institutional focus on online teaching would have. Thus, we wish to share our

experiences to regular faculty, who might want to add aspects of online teaching to their repertoire.

Table 1. Profile of participants for our training program.

	under-graduate	graduate	postdocs	faculty	total participants	female participants	total teams
Year 1	0	9	4	3	16	7	5
Year 2	0	14	2	1	17	6	5
Year 3	6	11	4	4	25	14	8
Total	6	34	10	8	58	27	18

Table 1 summarizes the basic profile of the participants for our program over the three years. We can see 1) most participants are graduate students since we believe graduate students are still in their early years of their research career and the offering of multidisciplinary education would have bigger impacts on their future career growth; 2) we try to address the under-representation of female researchers in STEM disciplines by having relatively equal number of female participants (27) and male participants (31). Some additional participants not eligible for NSF funding (not graduate students or post-docs/faculty) were included without support. An additional benefit for local participants was the three-credit special-topics graduate course.

3 REU Site: Online Interdisciplinary Big Data Analytics in Science and Engineering

We reported before in [4] that it is possible to involve undergraduate students in a program designed for advanced graduate students. In Year 3 of the Cyber-Training program, we were successful in applying for REU Supplement support for six undergraduate students at our institution from the NSF. We recruited for these positions in August 2019 and admitted two students from each discipline in September 2019. We report on how it is possible to successfully integrate undergraduate students in a program that was conceived for advanced graduate students and junior faculty. The key was to start the training for these local students during the fall 2019 semester. Since the students had a full course load to start with, the spreading out of material is crucial. The students were grouped by department during fall 2019 with a faculty mentor from that home department. They started by learning about the topics out of the 10 instructional modules that are in their own area, thus when they joined a multi-disciplinary team, they had all something to contribute. We then during winter 2019–20 leveraged the fact that the lecture videos of the first 10 modules are available for asynchronous delivery. The two teams of undergraduates in fact started on the homework and were able to get a head-start of several weeks of homework submissions before the official start of the program. Using the time thus freed up during several

weeks of instructions in Weeks 1 to 10, the undergraduate teams also started on research substantially earlier than Week 11. This concept worked and the undergraduate teams have results on the same level as the more senior teams.

Based on this experience, we applied and were awarded a summer undergraduate research program, called an REU (Research Experiences for Undergraduates) Site by the U.S. National Science Foundation (OAC–2050943) for programs in summers 2021, 2022, and 2023. The 2021 program involves 8 participants in 2 teams on the topics (i) Big Data and Machine Learning Techniques for Sea Ice Prediction and (ii) Big Data and Machine Learning Techniques for Medical Image Classification. The projects involve collaborators in the application areas, Yiyi Huang from NASA Langley Research Center and Jerimy Polf from the University of Maryland School of Medicine, respectively. The research uses the CPU, GPU, and Big Data clusters in the UMBC High Performance Computing Facility (HPCF, hpcf.umbc.edu), giving participants a real-life experience on a shared distributed-memory cluster running Linux with batch scheduling of jobs, etc. The teams are guided by the authors of this note as faculty mentors and by a dedicated graduate assistant. Each team ends the 8-week program with an oral presentation and a technical report published in the HPCF series. The participant teams additionally get exposure to a full set of professional development activities such as presentations by the Dean of the Graduate School and representatives from the Career Center.

This paper shares lessons learned of how to conduct a summer research program online for undergraduates from around the nation. We received more than 120 applications of highly qualified and motivated students from around the nation for the 8 NSF-funded slots. Each team is highly diverse in terms of race, ethnicity, gender, type of home institution, geographic home, and more. This provides challenges, but also incredible opportunities for professional and personal development of the participants.

Acknowledgment

This work is supported in part by the U.S. National Science Foundation under the REU Site (OAC–2050943), CyberTraining (OAC–1730250), and MRI (OAC–1726023) programs. The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF).

References

1. Computational Science: Ensuring America's Competitiveness. https://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf (2005)
2. National Academies of Sciences, E., Medicine, et al.: Future directions for NSF advanced computing infrastructure to support US science and engineering in 2017–2020 (2016)
3. Wang, J., Gobbert, M.K., Zhang, Z., Gangopadhyay, A., Page, G.G.: Multidisciplinary education on Big Data + HPC + Atmospheric Sciences. In: Proceedings

of the Workshop on Education for High-Performance Computing (EduHPC-17), p. 8 pages (2017). DOI 10.13016/M2KS6J78R

4. Wang, J., Gobbert, M.K., Zhang, Z., Gangopadhyay, A.: Team-based online multidisciplinary education on big data + high-performance computing + atmospheric sciences. In: The 16th International Conference on Frontiers in Education: Computer Science & Computer Engineering (FECS'20) (accepted (2020)). URL CyberTraining_FECS2020.pdf