Science DMZ and NSF CC* SWEETER Workshop Report

Organizers:

Dr. Smriti Bhatt and Dr. Izzat Alsmadi

Department of Computing and Cyber Security

Texas A&M University-San Antonio (TAMUSA)

Date, Time and Venue:

June 10, 2021

10:00 am – 12:00 pm (CT)

Virtual/Online at Texas A&M University-San Antonio

1. Executive Summary

The Science DMZ and NSF CC* SWEETER workshop, held virtually through zoom, on June 10, 2021, brought together a number of faculty and researchers, who are interested in or already doing data-intensive research and need large scale data computing or high performance computing (HPC) resources, from Texas A&M University-San Antonio (TAMUSA). Two research experts, Dr. Dhruva Chakravorty and Dr. Lisa Perez from Texas A&M High Performance Computing (HPRC) Center at Texas A&M University (TAMU), College Station, also attended the workshop. This workshop was organized as a collaborative effort by the National Science Foundation (NSF) CC* SWEETER grant (Award# 1925764) and the Science DMZ project at TAMUSA. The main objectives of the workshop were to (i) provide information on NSF CC* SWEETER grant and the Science DMZ project, (ii) initiate discussion on cyberinfrastructure, data computation and HPC needs and challenges at TAMUSA and how we can address them, and (iii) foster interdisciplinary research collaborations within TAMUSA and with other SWEETER partner and external institutions in the future.

Here, we present brief discussion on the two projects and their goals.

The **NSF CC* SWEETER: South West Expertise in Expanding Training, Education and Research** project is a multi-institutional partnership to promote regional Cyber expertise by leveraging resources and expertise from various partner institutions. SWEETER partners include Texas A&M University, University of Texas at Austin, two research-extensive institutions in the two major public university systems in Texas, joined by Prairie View A&M University, a historically black college or university (HBCU), and five Hispanic Serving Institutions (HSIs): New Mexico State University, University of Arizona, University of Texas at San Antonio, Texas A&M San Antonio and West Texas A&M University. The main goals of SWEETER are to enable peer-to-peer research support, organize on premise7 workshops, and build research collaborations across institutions by providing training and education opportunities. SWEETER unites multiple institutions ranging from not-for-profits, community colleges, minority serving institutions (MSI), research-intensive universities, and industry from different states to build a research network which consists of subject experts in large-scale data computing and shared HPC resources. Researchers at MSI and smaller institutions have limited Cyberinfrastructure and HPC resources which directly affect research excellence and efficiency at these institutions. SWEETER enables a diverse collaboration between research intensive larger institutions and smaller institutions and allows institutions to leverage expertise in different areas of computational sciences by employing a model of resource sharing where each institution is a provider and receiver of resources. The PI/Co-PI team includes -- Dhruva Chakravorty (PI, Texas A&M University), Timothy Cockerill (Co-PI, University of Texas at Austin), JoAnn Browning (Co-PI, University of Texas at San Antonio), Emily Hunt (Co-PI, West Texas A&M), Diana Dugas (Co-PI, New Mexico State University).

The **Science DMZ** project at TAMUSA is an initiative to build a campus wide research infrastructure at TAMUSA. A Science DMZ offers a network environment that is tailored to the needs of high-performance science applications, including large data transfers, remote

experiment control, and data visualization. The Science DMZ project proposes to develop a Science DMZ network at TAMUSA, which is distinct from the 'regular' network and addresses common network performance issues faced by researchers working with large amount of data (large data-intensive research), and is dedicated to researchers using advanced science applications. This new network would allow to cut data transfer times, enable seamless data collaboration and will set the stage for new discoveries on TAMUSA campus and beyond. The project aims to develop a dedicated Science DMZ of (40-100) Gbps connections across campus which would also be available to wider scientific community. Such cyberinfrastructure enables high bandwidth, storage and data transfer which are vital requirements to conduct impactful research in many research fields. They are essential for TAMUSA to support faculty and student research and open new internal and external collaboration opportunities.

A. Goals and Objectives

The overarching goal of this workshop is to facilitate development cyberinfrastructure at TAMUSA. In order to achieve this goal, we organized this workshop to bring together faculty and researchers working on data-intensive research and need HPC resources. The main objectives were to provide information on the National Science Foundation (NSF) CC* Team: SWEETER project which aims to provide education and training to researchers, and to discuss about the Science DMZ project at TAMUSA that will create necessary cyberinfrastructure for researchers. Another important objective was to initiate discussion on HPC and cyberinfrastructure needs of the HPC research community at TAMUSA and introduce available resources through external institutions/organizations, such as NSF XSEDE project (https://www.xsede.org/), Texas A&M High Performance Research Computing (HPRC) Center (https://hprc.tamu.edu/) and Texas Advanced Computing Center (TACC) at University of Texas at Austin (https://www.tacc.utexas.edu/).

B. Organization of the Workshop

The format of the workshop included dedicated presentation of 20-25 minutes on the two projects – NSF CC* SWEETER (by Dr. Smriti Bhatt, Dr. Dhruva Chakravorty, Dr. Lisa Perez) and Science DMZ at TAMUSA (by Dr. Izzat Alsmadi). We then had participant introduction and discussion of each participant's research areas and use cases/projects where they need large-scale data computation and HPC resources. This was followed by open discussion on research needs, external resources available through different programs, such as XSEDE, HPRC, public cloud, etc., to meet HPC and cyberinfrastructure needs, and what current challenges are being faced by researchers at TAMUSA impeding their progress.

There were also multiple short question and answer (Q&A) sessions during the discussion resulting in an interactive and engaging workshop. Based on the discussions with faculty on their research needs and challenges, there are different workshops and training sessions to be planned for the next academic year (as discussed below).

2. Open Discussion and Recommendations

After the presentations about the two projects, we had participant engagement and discussion on various topics including their research use cases/projects and need for HPC, current state and mechanisms of requesting and acquiring resources, and challenges and issues faced with discussion on workshop and training needed to overcome these challenges.

A. Needs and Challenges

Currently, TAMUSA does not have a HPC or large-scale cyberinfrastructure on campus. This is a critical issue for researchers whose research is focused on computational sciences and need HPC infrastructure and resources to conduct research experiments. Therefore, as a first step, we collaborated with the Texas A&M HPRC Center and became a partner institution in the NSF CC* SWEETER grant to facilitate research training, education, and enhance collaboration with larger institutions that have HPC resources. More recently, the Science DMZ project has also been initiated at TAMUSA to build a Science DMZ network infrastructure on campus for the researchers.

An important aspect of this workshop was to bring together these faculty/researchers and initiate a discussion on their needs and challenges, and how NSF CC* SWEETER and Science DMZ project can assist them in meeting their needs and overcoming current challenges. Faculty from various departments within different colleges attended the workshop and discussed about their research which requires large-scale data computation and their need for HPC resources.

To get access to HPC resources, some faculty reach out to external sources for getting resources, either by requesting free resources, or rent resources, or even outsource their computational tasks. One of such platforms is NSF XSEDE that provides computational resources to the community. For this, the researchers need to submit a proposal requesting specific resources along with a great detail of their projects and system configuration. The proposal may be accepted or denied after review. Other sources within the region are Texas Advanced Computing Center (TACC) at University of Texas at Austin, and Texas A&M High Performance Computing Research (HPRC) center at Texas A&M University, College Station. They have their specific processes for requesting advanced computation resources. The HPRC resources are available to TAMUSA researchers via collaborations with TAMU, or via a purchase of resources. Another option that researchers use for accomplishing their research tasks that need large-scale computation is Cloud Computing.

However, there are several challenges faced during the process of requesting these resources and also after they get access to resources. Some of the specific challenges that faculty shared during the workshop as outlined as follows.

- One of the main challenges that researchers face is lack of knowledge and information on local, regional, and national advance computing resources that are available to the research community.
- Researchers also face challenges while requesting access to computational resources from external institutions, such as the proposal and application process that require very specific details about the computational systems,

3

which the requestors are not aware of, or do not have enough knowledge about it. Faculty members shared their experiences with the XSEDE and their application process which requires specific details about the project and system resources. It is a challenge to provide enough details as a faculty/researcher may or may not be knowledgeable about it. Moreover, in some cases, they mentioned that their proposal was rejected due to lack of information. It is discouraging for researchers to go through a complicated process which may result in denied requests. Therefore, researchers need training on the proposal development and application process to successfully get access to the resources they need for their research.

- In some cases, the computation resources are approved for a limited amount of time that is not enough for the researchers to conduct their experiments. Besides, sometimes the systems are overcommitted and are very slow that it does not work at all and tasks are in pending state. This is not helpful to the researchers.

- For cloud computing resources, a major challenge is the cost. There is significant expense associated with cloud compute, storage, and network resources, such as advanced compute nodes, GPUs, large data storage, etc.

- Another issue mentioned with external resources is institution specific compliance policies and software licensing. Every institution managing HPC resources have specific policies that need to be satisfied for securely using the systems hosted in their network. A faculty member discussed about his experience on issues after getting access to the resources, such as software application compatibility and licensing issues.

- Furthermore, training on using HPC and advanced computing systems is also a challenge for faculty/researchers and students. For instance, they might not have all the knowledge on setting up the environment before submitting their tasks to the system. Training on programming (Python, R, etc.) for large-scale computing is also essential for faculty and students.

Due to these challenges, there was a discussion on the need to have on premise HPC resources. Researchers at TAMUSA have been submitting grants to build HPC infrastructure, such as a cluster on campus. We need to keep pursuing external grants from funding agencies (e.g., NSF) to build cyberinfrastructure at TAMUSA.

An initial step towards building cyberinfrastructure at TAMUSA is the Science DMZ project. Dr. Izzat Alsmadi, who is leading the Science DMZ project, also discussed about the need and impact of formal HPC education. He had recently taught a course related to HPC and recommended that we need to have such form of formal education for the research community.

B. Future Workshops and Trainings
As discussed above, there is a significant gap between researchers' needs and available resources, especially at smaller institutions as TAMUSA with limited to no HPC

resources. To overcome these challenges, we need to provide training and organize workshops for faculty and students based on their research needs and areas. SWEETER aims to bridge this gap by providing training (virtual and on-site) and organize workshops at different partner institutions. In addition, there is a network of subject experts, available through SWEETER, who can provide support in pursuing external funding (e.g., NSF Campus Cyberinfrastructure –CC*) to build cyberinfrastructure at TAMUSA. The Science DMZ project also has allocated funds to provide trainings for researchers and students.

Inspired by our open discussion session, we envision the following workshops and training programs in the future.

- A workshop focused on educating faculty/researchers about regional and national HPC and advance computing resources that are available for free or through subscriptions. It will also include a detailed session on how to submit strong proposal requests with required system specific information. This workshop addresses the needs and some of the challenges discussed in previous section.
- A hands-on training session on how to use advance computing resources once the access is grant to researchers. It would include basic environment setup, programming languages, and some example use cases.
- A training on programming languages for advance computing applications. It is not the basic programming languages, but is more focused on large scale data computation. The targeted audience includes faculty and students.

These are some specific workshops and training that faculty are interested in and requested. More faculty and student specific sessions can be organized to enhance the HPC research and education at TAMUSA.

3. Main Outcomes from the Workshop

This was the first workshop session organized for providing knowledge on SWEETER and Science DMZ projects, and to understand the needs and challenges faced by the research community at TAMUSA. The main outcomes of this workshop are an overview and evaluation of current state of cyberinfrastructure at TAMUSA, and a plan for the future to build necessary cyberinfrastructure at TAMUSA through projects like the Science DMZ as well as future external funding, and a strong research network and support available through the SWEETER and its partner institutions.

A. Evaluation of the Current State

It is a necessary task to get an overview of current situation and evaluate the current state of things, especially in the COVID-19 pandemic situation. Previously, we had collected science drivers from researchers for the SWEETER grant. However, due to COVID-19 pandemic, there was limited interaction with researchers in the last year, so it was essential to initiate discussion within the research community at TAMUSA and

learn about recent developments in research areas and use cases. More recently several use cases were requested from faculty/researchers by the Science DMZ team. This workshop played a vital role in bringing the HPC research community together at TAMUSA and initiate open discussion among them. In the future, it will also enable multi-disciplinary research collaborations across departments and institutions.

B. Plan for the Future

Based on interesting discussions among participants and their feedback to workshop organizers, the plan for the future is to organize and conduct local workshops and trainings for educating targeted audiences (faculty and/or students), provide information on workshops and trainings available through external sources, such as HPRC and TACC, provide collaboration and networking opportunities, and provide support for pursuing external funding to build and strengthen cyberinfrastructure at TAMUSA.

    i. **Workshops, Training and Education:** With improving COVID-19 pandemic situation, we plan to organize more in-person workshops and training programs for our researchers and invite subject experts from various SWEETER partner institutions including other institutions in the region. We have three major research institutions in the region, University of Texas at San Antonio (UTSA), Texas A&M University (TAMU), College Station, and University of Texas at Austin. Moreover, there are online trainings and workshops which will be made available to TAMUSA researchers (faculty and students). We also need to develop formal education pathways for HPC at TAMUSA.

    ii. **Collaboration and Networking:** This workshop was a step forward towards building new collaborations. In the future, we plan to focus on networking and open discussion sessions to foster collaborations between researchers internally and across the region. With the new Science DMZ at TAMUSA, we will have access to on-campus cyber resources for initiating collaborations with larger instructions and leverage their expertise and network. It is also necessary to include networking sessions in the future workshops to encourage collaborations.

    iii. **External Funding Opportunities:** SWEETER has several programs which focuses on providing support for smaller institutions in pursing external funding opportunities. During the workshop, researchers discussed their efforts for acquiring HPC resources on campus through external funding opportunities. Research intensive larger institutions have significant expertise in writing such grant proposals, thus we plan to leverage their expertise through the SWEETER network for upcoming funding opportunities. This is an excellent opportunity for developing collaborative proposals as well. An upcoming funding opportunity was also discussed during the workshop, the NSF Campus Cyberinfrastructure (CC*) solicitation which has a deadline of October 11, 2021. A team of researchers are interested and already initiated discussion on this solicitation with the

research experts from Texas A&M HPRC Center, Dr. Dhruva Chakravorty and Dr. Lisa Perez, during the workshop.

4. Acknowledgement

(Note: See Appendices in next page.)

5. Appendices
A. Participants

There were 11 participants who attended the workshop throughout the zoom session. There were 9 faculty members (including new faculty members in Biology department for fall 2021) from different departments and colleges at TAMUSA. Five faculty members from the Department of Computing and Cyber Security, two faculty members (including new faculty) from the Biology Department, one faculty from the Mathematics Department, and one from the Department of Accounting and Finance, and two participants (subject experts) from the Texas A&M University, College Station. A doodle poll was sent out to get the faculty/researchers' availability and accommodate their schedules.

This workshop was designed for faculty/researchers to provide them information about the NSF CC* Team: SWEETER grant and the Science DMZ project and to learn about their HPC and cyberinfrastructure needs and current challenges through focused open discussion and question and answer (Q&A) session.

B. Workshop Schedule and Agenda

The workshop schedule and agenda is as follows.

Activities	Time
• Welcome	10:00 am (5 minutes)
• Opening Remarks	10:05 am – 10:15 am (10 minutes)
• Science DMZ at TAMUSA: Introduction	10:15 am – 10:35 (20 minutes)
• NSF CC* SWEETER Project: Introduction	10:35 am – 10:50 am (15 minutes)
BREAK	10:50 am – 11:00 am (10 minutes)
• Researchers/participants Introduction and Project Discussion	11:00 am – 11:30 am (30 minutes)
• Open group discussions and Q&A: HPC resources, Science DMZ resources, and general guidance	11:30 am – 11:50 am (20 minutes)
• Closing remarks	11:50 am – 12:00 pm (10 minutes)