Unidata Community Equipment Awards Cover Sheet
UND Big Weather Web: Distributed Data Solutions

Date: Mar. 24, 2016

Principal Investigator: Gretchen Mullendore
Title: Associate Professor
Institution: University of North Dakota
Telephone number: 701-777-4707
FAX number: 701-777-5032
Street Address:
4149 University Avenue, Stop 9006
Grand Forks, ND 58202-9006

Email address: gretchen@atmos.und.edu

Institution Official: Barry Milavetz
Title: Associate Vice President for Research
Telephone number: 701-777-4280
Email add[redacted]

[redacted]
Dr. Gretchen Mullendore, Principal Investigator

[redacted]
Dr. Barry [redacted]

UND Big Weather Web: Distributed Data Solutions
May 15, 2016 to May 14, 2017
Proposed Budget: $12,582 (Total $24,973 with $12,391 contribution from UND)

1. Project Summary

This proposal will provide funding and support for testing of distributed data solutions using both a) local hardware and b) cloud services. Specifically, this proposal will integrate with the research and education initiatives already being conducted with support of the NSF Big Weather Web cyberinfrastructre grant. As part of this grant, forecast ensemble members are currently being created at several universities around the nation, including University of North Dakota (UND). These members are being uploaded to RAMADDA servers at each institution. UND is currently hosting RAMADDA on one of the computational clusters, which is not optimal. We propose to leverage the BWW grant to get partial funding from Unidata for the UND RAMADDA data server.

In addition to the local RAMADDA data server, we propose to utilize cloud services for both a second RAMADDA instance and to deploy WRF ensemble members in the cloud using WRF Docker. This component of the proposal will not require additional costs as our current plan is to request an educational allocation via Microsoft Azure. However, personnel support from Unidata will be needed. The PI (and the PI's institution) has no experience with cloud utilization for atmospheric research and educational purposes.

The major outcomes from this project are:
1. Integration of Unidata products and services into the UND's research and teaching activities.
2. Testing of two possible pathways for distributed data, specifically local computing and storage versus cloud computing and storage.

There has been a lot of debate on which method, local or cloud, is the better pathway for universities, who are often limited by traditional funding models. This project has the unique opportunity to actively test both pathways side-by-side and report to the broader community the advantages and disadvantages of each.

This project fits into two of the special consideration categories: remotely-accessible storage systems and remote server-based data analysis. This project will advance knowledge and understanding within the geosciences, both as a research project (contributions to the BWW research ensemble forecasting system), an education project (helping students learn more about ensemble systems and the tools used to run and analysis models), and as a technical project focused on testing different data sharing pathways in the academic workflow.

2. Project Description

2.1 Big Weather Web (BWW)

The Big Weather Web is a 3-year NSF project that started in August, 2015. The principal investigators/senior personnel include representatives from 8 universities, NCAR and Unidata (BWW, 2016). This proposal was motivated by several EarthCube workshops focused on the numerical weather prediction community (e.g., EarthCube, 2016). Specific problems impacting both research and education were identified via these workshops: poor reproducibility, lack of accessibility, and significant obstacles to the adoption of new technologies.

To address these identified problems, it was proposed to run a large ensemble forecasting system (EFS), distributed across 7 universities (UND, Texas Tech, SDSM, UW Milwaukee, U Albany, Penn State and CSU). The distributed EFS will be used to complete multiple research and education projects. The outcomes of using the EFS are to make high volumes of data available to participants and to enable participants to share tools, data processing platforms, and analysis techniques by utilizing common storage, linking, and cataloging methodologies. The goal is to break out of the "business as usual" mode of each university downloading 100% of the data to their local machines. Instead, data will be shared via RAMADDA servers. Five of the seven universities already have RAMADDA servers up and running (including UND); however, only two of the universities have the public sharing capabilities currently working. In addition to the distributed data access and analysis component, WRF Docker (Hacker et al., 2016) is being tested for easy ensemble member deployment and increased reproducibility of model results.

2.1.1 UND BWW Project

As mentioned above, the EFS will be used by individual investigators for a variety of projects. The project headed by UND (led by the PI of this equipment proposal, Dr. Mullendore) is an education integration project entitled "Numerical Weather Prediction Modules for Introductory and Advanced Undergraduate Classes". Specifically, we proposed creation of teaching modules for one undergraduate introduction class and one upper-division numerical methods class. At UND, the specific classes to be targeted are Meteorology I and Numerical Methods in Meteorology. Modules for both classes will include material for the classroom as well as a hands-on lab component. The modules will use the EFS and WRF Docker to convey concepts both applied, such as numerical forecast methodologies, and theoretical, such as numerical forecast uncertainty and nonlinear behavior (i.e. chaos). Although these modules are being developed for specific UND courses, they will be stand-alone modules that should translate to similar courses at other universities. Dissemination is planned via presentation at AMS Educational Symposia and via integration with existing User Workshops, such as the focus on observational data in geoscience education at the 2009 Unidata Triennial Users Workshop (Etherton et al. 2010).

2.2 Local Computing and Storage

The requested storage server has 44 TB of useable space. The server will provide file sharing capabilities to the system and host RAMADDA. This server will be connected to the department modeling/analysis clusters via NFS mount (simulations will be written to local disk during runtime and then transferred to the storage server). The physical connection is gigabit speed interconnect. As stated in the project summary, UND has a RAMADDA instance currently running on the simulation machine (a 24-processor Linux machine), which does not have sufficient disk capacity for long-term storage. For other projects, simulations have been archived in a piecemeal fashion, with some runs on small local external drives (2-3 TB each) or on older department storage drives. The existing department storage drives are all over 7 years old, and some have started to lose data and need to be retired. While the highest priority research/education project for the new storage server is the BWW project, we plan for this storage server to also eventually host other datasets, from multiple research groups, for dissemination via RAMADDA. For example, regional climate downscaling simulations are being run for the Northern Plains region, and this proposed server configuration would be an excellent method for sharing that dataset with partner institutions.

2.3 Cloud Computing and Storage

Allocations on Microsoft Azure will be requested via the Microsoft Educator Grant Program (Microsoft, 2016). Note that other Unidata cloud partners (e.g., Amazon) would also be acceptable; ultimately, the cloud service chosen would be based on availability of educator partnerships and recommendations of Unidata staff. Cloud allocations would be requested for the PI and for all students in the senior-level UND Numerical Methods in Meteorology class (estimated at 15 students). As the Azure awards are limited to class-related use, only the PI and students will be testing the local RAMADDA versus cloud RAMADDA access. However, students are great test subjects as they cover the spectrum from computer savvy to computer-phobic. It will be enlightening to see if their interaction with each system will be at all different. There will of course be differences from an administration standpoint; the PI will assess these differences.

In addition to hosting RAMADDA, students will also be running WRF instances in the cloud using WRF Docker. Some students will use the Azure allocation and some will be running WRF Docker on a local high performance cluster. As with the RAMADDA instance, our goal is to compare end-user experience and administrative overhead of the local computational clusters versus the cloud.

2.4 Project Benefits

The BWW infrastructure will allow students access to model ensemble members, and, more importantly, provide them with analysis tools and the workflow help to use those tools, so they can produce plots themselves. Most exciting for the students is the chance for students to also create *and disseminate* (via RAMADDA) their own runs through the Big Weather Web. Integration of current research with undergraduate education has been shown by many studies to be valuable for students (e.g., Mullendore and Tilley, 2014). In turn, working with students will benefit this proposed effort as the students will be introduced to the distributed workflow model, helping to widen adoption of this model.

2.5 References

BWW, cited 2016: Big Weather Web, NSF Award# 1450488 [Available online at http://bigweatherweb.org]

EarthCube, cited 2016: Executive Summary Workshop Results: Shaping the development of EarthCube to enable advances in data assimilation and ensemble prediction. [Available online at http://earthcube.org/document/2012/data-prediction-ensemble-assimilation-workshop-report]

Etherton, Brian J., Sean C. Arms, Larry D. Oolman, Gary M. Lackmann, and Mohan K. Ramamurthy, 2010: Using operational and experimental observations in geoscience education. *Bulletin of the American Meteorological Society*, 92(4):477–480, 2014/06/22 2010.

Hacker, J., J. Exby, N. Chartier, D. Gill, I. Jimenez, C. Maltzahn, and G. Mullendore: Collaborative research and education with numerical weather prediction in a common environment enabled by software containers. *96th AMS Annual Meeting: 32nd Conference on Environmental Information Processing Technologies*, 10-14 Jan, 2016, New Orleans, LA

Microsoft, cited 2016: Microsoft Azure in Education [Available online at https://azure.microsoft.com/en-us/community/education/]

Mullendore, G.L. and J. S. Tilley, 2014: Integration of undergraduate education and field campaigns: A case study from Deep Convective Clouds and Chemistry (DC3). *Bulletin of the American Meteorological Society*, 2014/06/22 2014.

3. Budget

3.1 Budget Justification

Included in the budget is $16,582 for a storage server, which includes $12,582 requested from Unidata and $4,000 provided by UND. Also included in the budget is salary for the PI, Gretchen Mullendore, for 0.45 months. This salary will cover integration of the data solutions into education and research activities. This salary is paid by UND. Fringe benefits are estimated at 30% for faculty/staff; actual fringe benefits will be charged to the project.

The University's approved indirect cost rate for research projects of 39.0% is included. It is based on modified total direct costs, which exclude equipment greater than $5,000.

We also expect a contribution via Microsoft Azure Cloud Services. We estimate an Azure award equivalent to about $12,000, consisting of one professor allocation (12 months @ $250/month) and 15 student allocations (6 months @ $100/month). The information provided is for proposal evaluation purposes only. The Microsoft Azure Cloud Services contribution will not be tracked and reported as cost share.

3.2 Line Item Budget

	Unidata	**UND**	**TOTAL**
Description:	Budget	Budget	Budget
Mullendore, Gretchen	-	4,644	4,644
Fringe Benefits (30% faculty/staff)	-	1,393	1,393
Total Personnel	-	6,037	6,037
Storage Server	12,582	4,000	16,582
Total Direct Costs	12,582	10,037	22,618
Indirect Costs/F & A	-	2,354	2,354
Total Aerospace Costs	**12,582**	12,391	24,973

*See next page for detailed summary of proposed storage server.

Base Unit	SCv2000 SAS (210-ADRS)	1
Base Unit	6Gb Mini-SAS to Mini-SAS Cable, 0.6M, Qty 2 (470-AAKJ)	1
	Thank you choosing Dell ProSupport. For tech support, visit http://www.dell.com/support or call 1-800- 945-335 (989-3439)	1
	Dell Hardware Limited Warranty Initial Year (997-8150)	1
	Dell Hardware Limited Warranty Extended Year(s) (997-8151)	1
	ProSupport: Next Business Day Onsite Service After Problem Diagnosis,Initial Year (997-8158)	1
	ProSupport: Next Business Day Onsite Service After Problem Diagnosis,4 Year Extended (997-8161)	1
	ProSupport: 7x24 HW / SW Tech Support and Assistance, 5 Year (997-8176)	1
	Hybrid Onsite/Remote Deployment of a Dell Storage SCv2XXX Series, up to 8 hosts (998-2207)	1
	US Order (332-1286)	1
	Dell SCv2000 Shipping (340-AMXR)	1
	12Gb SAS Dual Controller (403-BBIL)	1
	SCv2000/SCv2020 Bezel (325-BBLP)	1
	Dell 4TB, NLSAS, 6Gb,3.5",7.2K, HDD (400-AGDB)	12
	Rack Rails (770-BBOJ)	1
	SCv2000 Regulatory Label (389-BFTQ)	1
	Redundant Power Supply, 580W (450-AEIU)	1
	C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America (492-BBDH)	1
	C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America (492-BBDH)	1
	SCv20x0 SSN License (634-BCRD)	1
	SCv20x0 SCOS Base License (634-BCRE)	1

***Total Purchase Price:**	**$16,581.57**
Product Subtotal:	$16,581.57
Tax:	$0.00
Shipping & Handling:	$0.00
State Environmental Fee:	$0.00
Shipping Method:	LTL 5 DAY OR LESS

4. Project Milestones

4.1 Local Storage

The storage server equipment will be purchased and deployed as soon as possible after receipt of the award. A location has already been designated in the UND College of Aerospace Sciences server room for the storage server, if awarded. Linkages to existing computational clusters (currently running the UND BWW ensemble members) and deployment of the RAMADDA software on the new server should be accomplished within several weeks of the equipment arriving on campus.

4.2 Cloud Storage/Computing

It appears that one can apply for the Microsoft Educator Grant Program at any time. We would plan to apply for the PI access in May, with the student access to start in July (for UND Fall semester 2016 with some lead time for the graduate teaching assistants).