UND Big Weather Web: Distributed Data Solutions
2016 Unidata Equipment Award
Final Report

1. UND Proposed Work

UND proposed funding and support for testing of distributed data solutions using both a) local hardware and b) cloud services. Specifically, UND proposed to integrate with the research and education initiatives already being conducted under the NSF Big Weather Web (BWW) cyberinfrastructre grant. The BWW forecast ensemble members are currently being created at several universities around the nation, including University of North Dakota (UND). The funding from Unidata was proposed for a new storage server that would be optimized to host ensemble members using RAMADDA.

In addition to the local RAMADDA data server, UND proposed to utilize cloud services to deploy WRF ensemble members in the cloud using "WRF in a Box" (i.e., WRF Docker). This component of the proposal did not require additional costs as an educational allocation was requested. However, personnel support from Unidata was needed as the PI (and the PI's institution) had no experience with cloud utilization for atmospheric research and educational purposes.

The major proposed outcomes from this project were:
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.

2. Modifications to UND Proposed Work

Subsequent to the award, three major changes occurred that modified the scope of work:
1. The RAMADDA architecture was deemed as the wrong software for the BWW grant. BWW is now hosting data using THREDDS from a single unified location (i.e., university members are not running THREDDS individually). Instead, the new storage server is running as a local backup to the UND members being pushed to the cloud storage (and served via THREDDS).
2. Microsoft Azure was investigated, but deemed to launch too slowly for classroom use. Instead, Amazon Web Services (AWS) was used, with free credits obtained via AWS Educate (https://aws.amazon.com/education/awseducate/).
3. Due to time constraints regarding creation of the software containers and materials for teaching, as well as some security concerns about installing Docker on local lab machines, only the use of the cloud computing pathway was investigated in the classroom setting. However, the usage of the local computational pathway versus cloud computing was investigated in a separate student project.

3. UND Completed Work

3.1 Storage Server Purchase
A 48 TB SAS storage server was purchased using combined funding from the 2016 Unidata Equipment Award and UND. This server is available via NFS mount to dedicated analysis servers, model simulation servers, and student desktop machines. Forecast outputs of UND's BWW ensemble members are stored on the newly purchased storage server. As noted in Section 2, the original plan of serving data via RAMADDA from this server was modified; currently model output is available only via traditional Linux mount pathways. The new storage server will continue to store existing UND BWW ensemble members, as well as new members, which are still being created in near real time. In upcoming semesters, the ensemble members on the storage server will be made available to undergraduate classes for comparing BWW ensemble members to the members being generated in the classroom.

3.2 Classroom Implementation

Creation and usage of a classroom module was completed for the Fall 2016 Numerical Methods for Meteorologists course at UND (See et al. 2017). Through the use software containers, a portable numerical weather prediction model was created at NCAR ("WRF in a Box"; Hacker et al. 2016). Separate containers were created for the WRF model, NCL post processing software, and GFS/GEFS input data. These containers were then accessed through Amazon Web Service's (AWS) Infrastructure as a Service (IaaS). Early discussions with Unidata staff were very helpful for both choosing the "best" cloud provider for this implementation, and also getting initial cloud instances running and configured. Using AWS and previously mentioned software containers, each of the 13 undergraduate students completed a homework assignment. This work required creation of a personal cloud computing instance, installation of the Docker software container application within the instance, and completion of 3 different WRF model runs. The homework assignment was used to assist in teaching of model uncertainty, parameterizations, and ensemble forecasting.

For this specific module implementation, the student feedback highlighted challenges relating to initial connection to AWS, failure of the students to understand the difference between the software containers and the cloud computing instance, and difficulty in understanding new software syntax. These difficulties arose from the way the information was relayed to students. A better introduction of cloud computing and containerization is needed for future classroom offerings. Despite these challenges, the module implementation was deemed a success. Every student completed the entire assignment and had a relatively easy time accessing AWS after initial issues were resolved. The student feedback was very positive regarding their perceived value in getting hands on experience using numerical weather prediction models. The module will be used again at UND in Fall 2017, and plans are to make the teaching materials available to other universities.

3.3 Economic Analysis

To complete model runs students used AWS's Elastic Compute Cloud (EC2). Students used general compute instances known as t2.large, which are charged on a per hour basis ($0.094 per hour uptime) and partitioned 60 GB of Elastic Block Storage (EBS). EBS charges $0.1/(GB*Month) and is prorated to the nearest hour of usage. The classroom usage occurred over a two week period and through the usage of AWS t2.large instances and EBS the module cost totaled **$70.98**. Breaking down the costs to each AWS feature, EC2 totaled $40.21 and EBS $30.77. The high cost of EC2 is attributed to improper shut down of user instances, which accrued additional costs overnight while instances were not in use. Proper instance management is necessary for economic efficiency within a classroom environment. A majority of these costs were covered through AWS Educate credits, which are offered to students and professors.

An undergraduate student focused their senior capstone project on comparing cloud computing and local computing. They compared the cloud computing scenario to use of local UND storage (as purchased through this award) and compute servers. The local compute server had 24 cores, 80GB RAM, and 14TB storage for $15,000. While this was a one-time purchase (and could be depreciated over several years), the additional costs such as employee salary and electricity quickly added several thousand dollars to the total cost estimate. At the time of this cloud-to-local comparison, there were no 24 core machines available through AWS, and the closest in comparison was the 16 core m4.4xlarge. The m4.4xlarge was charged hourly at $0.796. The undergraduate student tested 3 scenarios, two 3 hour runs a day, four 3 hour runs a day, and running 24 hours a day. Priced out to a cost per year the totals came to $1,719.36, $3,438.72, and $6,877.44 respectively. Note that long-term storage of data would also increase cloud computing costs. In summary, a simplified cost comparison for real-time forecasts showed comparable yearly costs for local ($6,500) and AWS ($7,000) computing pathways. It should be noted, however, that each use case is very sensitive to workflow decisions (e.g., see presentations from Unidata 2017), and more economic analysis is needed.

4. Conclusions

Traditional and cloud computing have their strengths and weaknesses. Cloud computing offers access to increased computing power at a reasonable cost. Computing costs can be monitored and manipulated to reduce costs based on computational need, urgency of completion, and optimized storage. Cloud computing is an increasingly popular option within the atmospheric sciences private sector. Increasing student exposure to a platform used by businesses is beneficial for work force development. Having exposure to cloud computing and software containers as both an undergrad and a graduate student is an advantage when seeking employment. A drawback of cloud computing is that it can be quite costly. If there is improper management of instances the cost of cloud computing usage can be prohibitive. Cloud computing can quickly drain the funds set aside, leaving a project without any compute power available to complete the tasks proposed. Additionally, when incorporating cloud computing into the educational environment, the learning curve associated with new software can be a challenge for some students.

5. References

Hacker, J., J. Exby, D. Gill, I. Jimenez, C. Maltzahn, T. See, G. Mullendore, K. Fossell, 2016: A containerized mesoscale model and analysis toolkit to accelerate classroom learning, collaborative research, and uncertainty quantification. *Bull. Amer. Meteor. Soc.,* **98**, 1129–1138, https://doi.org/10.1175/BAMS-D-15-00255.1.

See, T. W., G. L. Mullendore, J. Exby, and K. Fossell: Improving Model Accessibility for Undergraduate Students. 26th Symposium on Education, AMS Annual Meeting, 22-26 January 2017, Seattle, WA

Unidata, 2017: 2017 Modeling Research in the Cloud Workshop Agenda, https://www.unidata.ucar.edu/events/2017CloudModelingWorkshop/#schedule