UNIDATA COMMUNITY EQUIPMENT AWARD PROPOSAL

Enhancing Undergraduate Python and Modeling Skills: A Jupyter Notebook Multi-Core Server at Central Michigan University

Date: 3-28-17

Principal Investigator:	Dr. John T. Allen
Title:	Assistant Professor of Meteorology
Institution:	Department of Earth and Atmospheric Sciences, Central Michigan University
Telephone:	989-774-1923
Fax:	989-774-2142
Street Address:	Brooks Hall 326
	Central Michigan University, Mount Pleasant, MI, 48858
Email:	allen4jt@cmich.edu
Signature:	

Co-Principal Investigator:	Dr. Martin Baxter
Title:	Associate Professor of Meteorology
Institution:	Department of Earth and Atmospheric Sciences, Central Michigan University

University Official:	Dr. Lawrence Lemke
Title:	Professor & Chair, Department of Earth and Atmospheric Sciences
Telephone:	989-774-1165
Email:	lemke1ld@cmich.edu
Signature:	

University Official:	Dr. David Ash
Title:	Vice President for Research and Dean of Graduate Studies
Telephone:	989-774-3094
Email:	ash1de@cmich.edu
Signature:	

Requested Amount: **$15,046.26**

Enhancing Undergraduate Python and Modeling Skills: A Jupyter Notebook Multi-Core Server at Central Michigan University

John T. Allen and Marty Baxter

1. Project Summary

The meteorology program within the Department of Earth and Atmospheric Sciences at Central Michigan University (CMU) is requesting support to replace a failing server used for teaching with a standalone multi-core server for Jupyter Notebook development, Python instruction, and introduction to atmospheric modeling. This piece of hardware is important to the program as it provides an interface for students to interact with AWIPS-II data via an existing EDEX server, and allows students to work with larger remotely sourced datasets than possible on the existing 24 machines in the Meteorology Computer Lab. This server will be used to enhance student experiences and preparation for future careers by further integrating programming skills in Unix and Python into courses throughout the meteorology program, along with developing key atmospheric modeling and visualization skills. In addition to the local benefits of this server for students at CMU, this server will provide opportunities for continual development of teaching tools and Jupyter notebooks that will be shared with the wider Unidata research and educational community via the RAMMADA server and Github.

2. Project Description

The meteorology program at CMU has been an active member of the Unidata community since the mid-1980s, and has been a member of UCAR since 2007. Contributions to the Unidata community by meteorology faculty include: service on User and Strategic Advisory Committees, attendance at Unidata workshops, teaching students to use Unidata software, and sharing our expertise to others in the Unidata community. To better allow CMU faculty and students to take advantage of the latest tools and data Unidata has made available, we request funds for a standalone multi-core server, which will be dedicated to student learning opportunities in:

- Manipulating and displaying data in Python using Jupyter notebooks and the Bash shell.
- Accessing remote large datasets for case studies in the Synoptic Meteorology, Mesoscale Meteorology and Atmospheric Modeling classes.
- Utilizing AWIPS-II data retrieved via an existing EDEX server for educational purposes.
- Performing simple model runs using the Weather Research and Forecasting Model (WRF) and the Community Earth System Model version 1.2 (CESM) as part of the senior undergraduate Atmospheric Modeling course.
- Conducting undergraduate research projects which involve the analysis of reanalysis, observations and numerical weather model data, with faculty supervision.

This server is planned to be accessible both internally from the existing Meteorology Lab, and externally via virtual clients provided to students using Cisco AnyConnect.

As the faculty assist students in achieving the above outcomes, we will develop materials that allow others in the community to leverage our capabilities. Specifically, teaching materials and tutorials that make use of Python to analyze and visualize data will be developed using Jupyter notebooks and videos, which will be made accessible to the wider Unidata community via our RAMMADA server and Github. Examples of code already developed for use by students in our program include fundamental skills in atmospheric modeling (running and modifying code to produce output, and subsequently plotting an evolution of the 1-D wave equation using a script that generates the output as MP4 movies to visualize the instabilities of differencing approaches), and Python techniques for retrieving and displaying numerical weather model output data along with simple array manipulation.

2.1 Equipment Requested

Waterspout.eas.cmich.edu was acquired with a Unidata equipment proposal in 2004, and served until 2010 when it became unstable and overloaded. A replacement machine (also called *waterspout)* was purchased via a Unidata equipment proposal in 2010, and performed a variety of tasks as a RAMADDA server and a place for students to store data for research and class projects. Recently, significant instability has occurred on the system, requiring a complete reinstall and reboot, which resulted in renaming of the system to *foehn*. While *foehn* is still operable, the machine is now out of warranty and showing increasing instability. This suggests a likelihood of failure in the next year or two, which would require replacement of components of the hardware. This server is a critical part of the computer infrastructure for the Meteorology program at CMU, acting as a RAMADDA server for IDV case studies in Synoptic Meteorology, and as a place for students to run the WRF model as part of an Atmospheric Modeling class project.

The system acquired using the requested funds will be a Dell Poweredge R730 server, with the following configuration:

- Intel Xeon E5-2690 v4 2.6Ghz processor with 35M Cache and dual 14 core sets for a total of 28 usable cores.
- 12x16GB RDIMM, 2400 MT/s RAM for a system total of 192GB.
- PERC H730P Raid Controller
- Root file system and scratch space of 2x480GB Solid State SAS Drives for read intensive usage.
- Storage of 6x1.8Tb 10K RPM Nearline SAS drives.
- Intel Ethernet X540 10Gb/s ethernet.
- 5 Year Basic Hardware Warranty.
- Centos OS-6.

The march of technology has meant that the system requested is considerably more powerful than the original *waterspout/foehn,* which only had 12 cores in total. This proposed system will allow multiple students to run Python scripts or simple idealized or small scale simulations (WRF, CESM 1.2) in Mesoscale and Atmospheric modeling classes simultaneously. This is important given graduating classes of 12 or more students, and for engaging in programming instruction throughout the program which currently contains approximately 60 majors. As there will be considerable writing and re-writing on the hard disks, the storage redundancy will be configured using RAID 10 architecture.

2.2 How the Equipment will meet the Goals of the Project

The meteorology program at Central Michigan University is undergraduate focused, with approximately 60 signed majors and graduating classes between 10 and 15 students. With this focus comes a unique opportunity to develop students' programming skills, which will allow them to more easily transition to graduate school or the workplace. The currently required programming course for the meteorology program is taught by the computer science department using Java. This course is becoming increasingly obsolete for meteorological graduates. With the increasing reliance on large datasets and demand for computational skills in meteorology graduates, there is a need to rethink how we go about integrating programming specifically applied to meteorological problems into the curriculum. Python has become one of the most frequently used languages in the Atmospheric Sciences, and is relatively easy for students to become comfortable with. Despite its relative simplicity, Python easily interfaces with a wide range of data formats, including those that can make use of OpenDAP and other web-accessible protocols. Jupyter notebooks provide an interactive method for instruction, interspersing code with comments and instructions, along with opportunities to make running or modifying the code interactive. The faculty intend to integrate Python into the meteorology program via simple tasks for visualizing data in the first courses in the sequence, with more advanced data analysis and visualization as students move through the program (**Figure 1**).

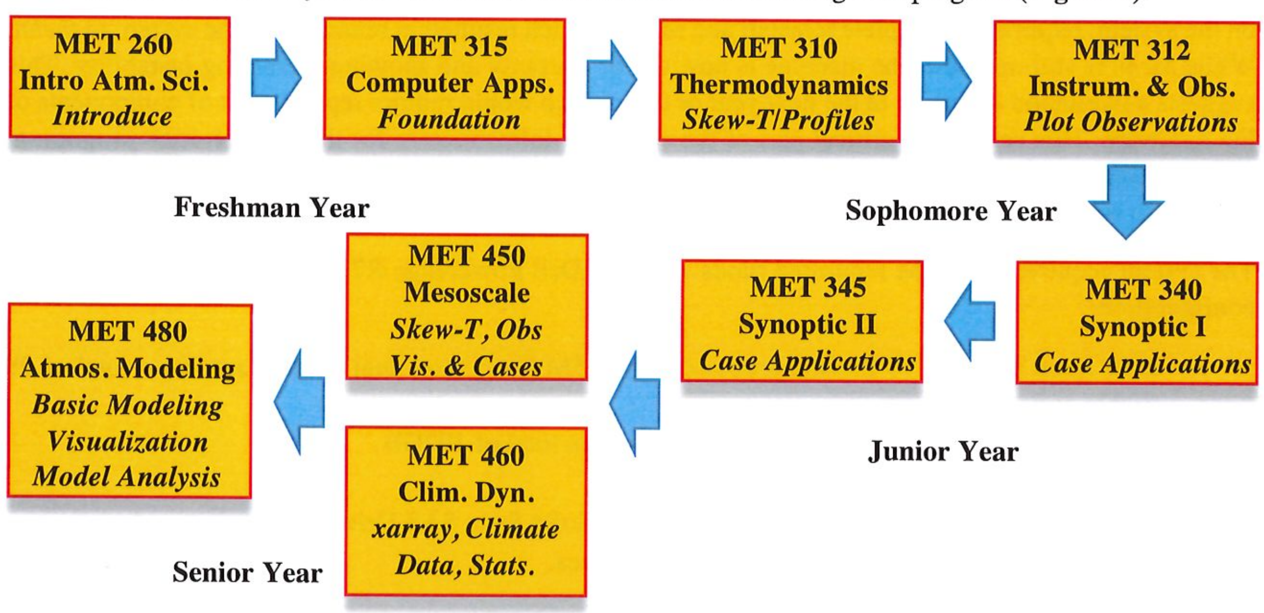

Figure 1: Planned integration of Jupyter notebook based Python learning for selected subjects the CMU Meteorology Program.

A student's initial introduction to programming will be undertaken in the first course of the major sequence, Introduction to Atmospheric Science. Exercises in this class will mostly involve demonstrations of how to display different types of meteorological data. Following this softer introduction, Computer Applications in Meteorology will then act as a foundation for the basic reading and manipulation of data, display and visualization techniques using software such as *MetPy*. These skills are built on by visualizing Skew-T diagrams and calculating parameters relevant to the thermodynamic state of the atmosphere, which combines well with learning the concepts behind the thermodynamic diagram in the sophomore-level course, Atmospheric Thermodynamics and the Boundary Layer. Another sophomore-level course, Meteorological Instruments and Observations, provides an ideal opportunity to learn how to use Python to analyze and display data obtained via the EDEX server. In junior-level Synoptic Meteorology (I & II), students will explore cases stored on the RAMADDA server, and use basic skills in Python to explore case study data and visualize the concepts they are trying to understanding in the coursework. In the fall semester of the student's senior year, they take both Global Climate Dynamics and Mesoscale Meteorology. The first of these two subjects provides ample opportunity to use the Python module *xarray* to explore larger climatological datasets through plotting and statistical analysis of trends and changes in atmospheric variables. In Mesoscale Meteorology, students will use Python to explore Skew-T diagrams in more detail, to calculate parameters for severe thunderstorm favorable environments, and to combine the host of skills developed throughout their earlier years to prepare an analysis of observational and model data for an end of semester scientific presentation of a case study. Finally, leveraging all these skills in the final course in the program, Atmospheric Modeling, students will code and run simple models in Python and Fortran. They will also develop skills to visualize data that they retrieve via Bash commands (such as GFS operational output at 3-hourly timesteps), and run simple model simulations in WRF and CESM1.2, using Python scripts to visualize the output and verify the forecasts for a chosen case.

2.3 Benefits to research, education and the Unidata community

In addition to the direct benefits to CMU students, the establishment of the Jupyter notebook server and undergraduate modeling capability will provide a code and instructional resource for faculty interested in taking a more active approach into integrating Python programming and visualization. The meteorology faculty at CMU are active and heavy python users for research purposes (as a scripting language, for visualization, and as way to improve the user-friendliness of legacy code in Fortran) and are actively working to enhance the development of visualization and data manipulation skills in our undergraduates, skills which have become increasing important in subsequent student career paths. One of the challenges facing instructors and student learners in the field is the wide array of packages available that are capable of completing many tasks, and a relatively small number of structured examples available that apply these different individual modules to common meteorological problems. Python provides a user friendly interface for students to obtain and visualize meteorological data in order to develop their understanding of meteorological concepts. Jupyter notebooks have been shown to be very effective for teaching students how to employ Python to investigate a dataset by providing comments and interactive code (Shen 2014), and they will thus be used throughout our courses. In addition, instructional videos have also been widely and successfully employed in the community (for example CESM training workshops). Thus, videos will also be developed for both elementary and specialized module usage in instances where students might struggle with learning from the notebook structure. These materials are already being developed in existing courses such as Computer Applications for Meteorology, Mesoscale Meteorology, and Atmospheric Modeling, and will be further developed over the next two years as approaches are tested and evolve.

A further benefit of improving student familiarity with programming is that this skill will provide students with a stronger foundation for pursuing research with faculty, or at summer internships and research experiences for undergraduates. The Jupyter notebooks and instructional videos should also yield similar benefits for other faculty instructing undergraduate and graduate students who are unfamiliar with coding in Python, further promoting an improved research culture under a single common programming language.

2.4 Relationship of proposed hardware to Existing Facilities and Overall Planning

CMU has a Meteorology Computer Lab that contains 24 student workstations that run often-used meteorological software, such as IDV, AWIPS-2, GEMPAK, GrADS, Gr2Analyst, ArcGIS, and the

Anaconda distribution of Python 2.7 (**Figure 2**). The lab also features an 88" Smartboard, traditional projector, and 16 screen electronic map wall. Two screens in the hallway outside the lab display output from our Campbell Scientific Weather Station. All of these resources were purchased with university funds.

The program maintains two servers in addition to *foehn,* both of which were purchased with university funds. Thus, the requested server will comprise one-third of the program's servers. *Vortex* was purchased in mid-2016, and serves as our EDEX server. *Weather* was purchased in mid-2015. It is used to run a real-time 4-member, 4 km WRF ensemble over the Great Lakes using the Unified Environmental Modeling System. The output from the model is used to create Campus Weather Forecasts in the Synoptic Meteorology sequence. The three servers will allow the program to maintain a robust set of activities, with each activity taking place on its own designated server: real-time data gathering and serving via EDEX (*vortex*), real-time numerical modeling (*weather*), and data analysis, visualization, and simulation (*foehn*), as shown in **Figure 3**.

Figure 2: Central Michigan University Meteorology Computer Lab.

Meteorology faculty have access to two local research clusters obtained with startup money, in addition to buy-in resources on Michigan State University's high-performance computing clusters. The requested server will be dedicated to undergraduate student teaching and research. While the machine requested is relatively small by most standards, it will have a massive impact on student engagement in programming tasks and developing rudimentary skills in atmospheric modeling. Our view is that the requested server will help provide for an important evolution of the meteorological program at Central Michigan University and supply many years of stable operation and engaging learning experiences in the years to come.

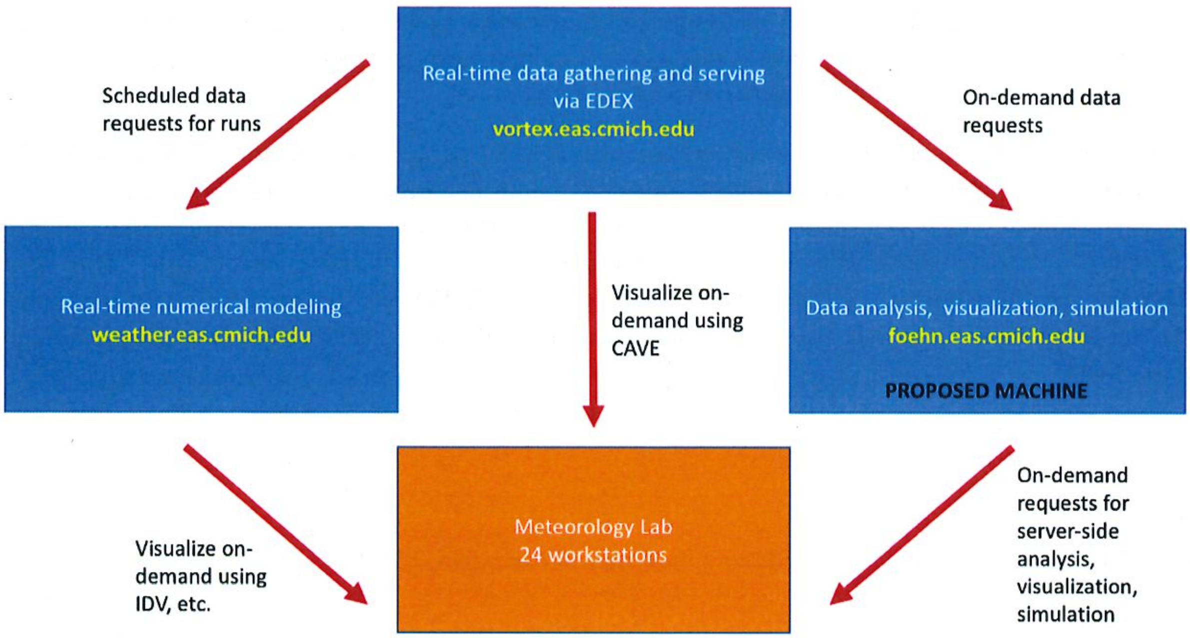

Figure 3: Architecture of existing meteorology computing facilities in the Earth and Atmospheric Sciences Department at Central Michigan University. Blue boxes indicate servers, orange box indicates clients, and red arrows indicate flow of data.

3. Budget

This proposal is for hardware only, with a full waiver of indirect cost. The development of Jupyter notebooks and video tutorials will take place as part of University supported teaching commitments. Our system administrator will install and configure all hardware and required OS software. The quote for this hardware purchase comes from Dell (see attached), and includes the educational premier discount. Cost sharing is not part of this proposal, however the Department of Earth and Atmospheric Sciences and College of Science and Engineering has already invested significantly in hardware to support the program, purchasing workstations (x24), an EDEX server for AWIPS-II, a machine for running a real-time forecast model, an electronic mapwall and other observational facilities. The quote for the server described in Section 2.1 is $15046.26, including shipping.

Total Cost: $15046.26

4. Project Milestones

A quote has already been obtained for this server from Dell (see attached). It is anticipated that hardware for the server will be purchased immediately following the receipt of funding, to allow installation of equipment as early as possible during the summer break so that all systems are configured and operational prior to the beginning of the 2017-2018 academic year. As there is existing space available in the centralized CMU server farm with integrated rack space there are no installation issues that will alter the planned installation goal. Installation, preliminary setup and ongoing support will be handled by the College of

Science and Engineering IT administrator designated to the Department of Earth and Atmospheric Sciences, who has significant experience in the configuration of new servers on behalf of the department. There is no cost associated with hosting the server in this facility, and the room has inbuilt power supply redundancy, climate control and monitoring to ensure ongoing operation.

The initial development of online training and coding materials including Jupyter notebooks and associated video tutorials for the more difficult concepts will take place in a progressive fashion over the next year, with development continuing indefinitely. This will allow for improvements based on student feedback, resulting in student-focused exercises that foster the most useful learning experiences. The Jupyter notebooks and video tutorials will be hosted online via RAMMADA and Github indefinitely, and updated or added to as new or relevant Python modules become available or are found to be relevant for career development.

References:

Shen, 2014: Interactive Notebooks: Sharing the Code. *Nature,* **515**, 151-152.

eQuote Details

Description	Quantity	Price
pe_r730_1356 | PowerEdge R730	1	$24,666.00
Premier Discount		-$9,619.74
		$15,046.26

Module	Description	Product Code	Sku	ID
PowerEdge R730	PowerEdge R730 Server	R73X	[210-ACXU] [329-BCZK]	1
Trusted Platform Module (TPM)	No Trusted Platform Module	NTPM	[461-AADZ]	1574
Chassis Configuration	Chassis with up to 8, 2.5" Hard Drives	258HD	[350-BBEN]	1530
Shipping	PowerEdge R730 Shipping	DSHIP	[340-AKKB]	1500
Processor	Intel® Xeon® E5-2690 v4 2.6GHz,35M Cache,9.60GT/s QPI,Turbo,HT,14C/28T (135W) Max Mem 2400MHz	26900	[338-BJCT]	1550
Additional Processor	Intel® Xeon® E5-2690 v4 2.6GHz,35M Cache,9.60GT/s QPI,Turbo,HT,14C/28T (135W) Max Mem 2400MHz	E52690	[338-BJEH]	1551
Processor Thermal Configuration	2 CPU Standard	2CPU	[370-ABWE] [374-BBHM] [374-BBHM]	1697
Memory DIMM Type and Speed	2400MT/s RDIMMs	2400MT	[370-ACPH]	1561
Memory Configuration Type	Performance Optimized	PEOPT	[370-AAIP]	1562
Memory Capacity	(12) 16GB RDIMM, 2400MT/s, Dual Rank, x8 Data Width	16GBMM	[370-ACNX]	1560
RAID Configuration	No RAID for H330/H730/H730P (1-16 HDDs or SSDs)	NRH330	[780-BBJS]	1540
RAID Controller	PERC H730P RAID Controller, 2GB NV Cache	H7302G	[405-AAEH]	1541
Hard Drives	(2) 480GB Solid State Drive SAS Read Intensive MLC 12Gbps 2.5in Hot-plug Drive, PX04SR (6) 1.8TB 10K RPM SAS 12Gbps 512e 2.5in Hot-plug Hard Drive	48RMLC 18TB15	[400-AMBW] [400-AJQV]	1570
Network Daughter Card	Intel Ethernet X540 10Gb BT DP + I350 1Gb BT DP Network Daughter Card	X540DC	[540-BBCC]	1518
Embedded Systems Management	iDRAC8 Express, Integrated Dell Remote Access Controller, Express	I8EXP	[385-BBHN]	1520
Internal Optical Drive	No Internal Optical Drive	NODVD	[429-AAOJ]	1600
Bezel	Bezel	BEZEL	[350-BBEJ]	1532
Rack Rails	ReadyRails™ Sliding Rails Without Cable Management Arm	RNOCMA	[770-BBBQ]	1610
Power Management BIOS Settings	Performance BIOS Setting	HPBIOS	[384-BBBL]	1533
Power Supply	Dual, Hot-plug, Redundant Power Supply (1+1), 1100W	1100R	[450-ADWM]	1620
Power Cords	(2) NEMA 5-15P to C13 Wall Plug, 125 Volt, 15 AMP, 10 Feet (3m), Power Cord, North America	125V10	[450-AALV]	1621
System Documentation	No Systems Documentation, No OpenManage DVD Kit	NODOCS	[631-AACK]	1590
Operating System	No Operating System	NOOS	[619-ABVR]	1650
OS Media Kits	No Media Required	NOMED	[421-5736]	1652
Order Information	US No Canada Ship Charge	USNONE	[332-1286]	111
PCIe Riser	Risers with up to 1, x8 PCIe Slots + 2, x16 PCIe Slots	RSR12	[330-BBCO] [330-BBCP] [374-BBHS]	1510

Hardware Support Services	5Yr Basic HW Warranty Repair, 5x10 NBD Onsite	U5OS	[976-8706] [976-8795] [991-2878] [996-8029]	29
Deployment Services	No Installation	NOINSTL	[900-9997]	714
Remote Consulting Services	Declined Remote Consulting Service	NORCS	[973-2426]	735

eQuote Subtotal	$15,046.26
Shipping*	$0.00
Shipping Discount*	$0.00
Tax*	$0.00
Environmental Disposal Fee*	$0.00

eQuote Total* **$15,046.26**