WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

UW Reference #MSN217577

University Corporation for Atmospheric
National Science Foundation

PI: Gregory Tripoli

A community THREDDS/ADDE data server and IDD infrastructure upgrade at UW-Madison

This proposal has been administratively approved on behalf of the Board of Regents of the University of Wisconsin System and is submitted for your consideration. Please keep our office advised as developments occur with regard to this application.

The appropriate programmatic and administrative personnel of each institution involved in this grant application are aware of the sponsor's grant policy and are prepared to establish the necessary inter-institutional agreement(s) consistent with that policy.

All costs cited conform to established institutional policies and procedures. Our DHHS Negotiated Rate Agreement can be found at http://www.rsp.wisc.edu/rates/rates.pdf. Website: http://www.rsp.wisc.edu/

A final agreement is contingent upon the successful negotiation of terms and conditions acceptable to the University of Wisconsin-Madison.

We ask that you use the University's above-referenced proposal number in any future correspondence.

Questions regarding administrative matters should be directed to:

PreAward Services by email: preaward@rsp.wisc.edu or by phone: (608) 262-3822

Questions regarding the technical nature of this application should be directed to:



Managing Officer

March 15, 2018

Office of Research and Sponsored Programs
University of Wisconsin-Madison 21 N. Park St, STE 6401 Madison, Wisconsin 53715
608/262-3822 Fax: 608/262-5111 www.rsp.wisc.edu

Unidata Community Equipment Awards
Cover Sheet

A community THREDDS/ADDE data server and IDD infrastructure upgrade at UW-Madison

Date: March 14, 2018

Principal Investigator Name: Gregory Tripoli
Title: Department Chair
Department: Department of Atmospheric Sciences – University of Wisconsin – Madison
Telephone Number: (608) 262-9845
FAX Number: (608) 262-0166
Street Address: 1225 W. Dayton St, Madison, WI 53706

Email address: tripoli@aos.wisc.edu

Signature of PI [redacted]

Name of Institution Official: Brenda Egan
Title: Managing Officer
Institution: The Board of Regents of the University of Wisconsin System

Street Address:
University of Wisconsin-Madison
Research and Sponsored Programs
21 N. Park Street, Suite 6401
Madison, WI 53715-1218

Telephone Number: (608) 262-3822
FAX Number: (608) 262-5111

Email address: preaward@rsp.wisc.edu

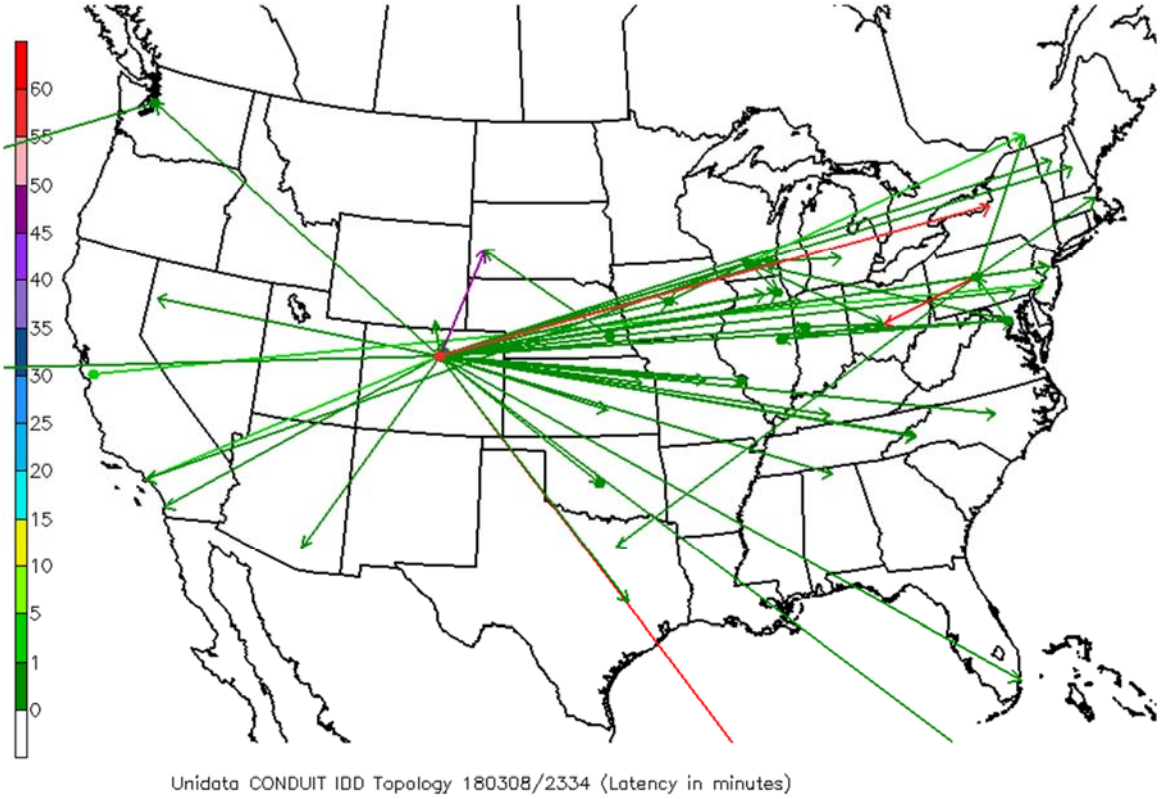
Signature of University Of [redacted]

Project Summary

Since the early days of the Unidata Internet Data Distribution project (IDD), the UW-Madison Atmospheric and Oceanic Sciences department has served as a top-level node for data distribution, and has maintained a local data server used in-house with GEMPAK, the Unidata IDV, etc. Data saved at UW-Madison has frequently been offered to other participating UNIDATA sites to fill in data gaps or help with research or case study data needs.

Over the past several years, the amount of observation and model data available for weather and climate systems has increased dramatically. Model forecasts from the RUC, HRRR, NAM, GFS and GEFS, among others, are available at increasingly high time and spatial resolutions. Similarly, the higher time and space resolution, as well as a tripling of the frequency bands available from the GOES-R series of satellites are generating an order of magnitude increase of satellite data alone.

At the same time, there has been a trend away from the more tree-based IDD of the past, to a more hub/spoke based IDD, where the majority of IDD sites are pulling data from Unidata directly, and not relaying to other sites. Figure 1 shows a stark example of this using CONDUIT data from March 8, 2018.



(Figure 1 – Conduit topology for March 8, 2018)

Part of this trend is likely due to staff with knowledge of the ldm/idd retiring or moving on, and departments not having someone to manage local data servers and relays, or not seeing data ingest and relay as a high priority to the department's mission. Another part is likely due to the sheer amount of data and the size of it all, much of which, while convenient to have available, might never be used, and is likely available on one or more remotely accessible sites.

Earth science data sets have grown in size and complexity to the point where it is neither feasible nor desirable to send all of the data from all sources via the internet to all downstream sites. Instead, remote data cataloging and access utilities such as THREDDS and RAMADDA have been developed. These technologies allow for large data sets to be housed on remote servers and accessed on demand when needed by software such as the Unidata IDV or through an interface like Unidata's siphon package, which enables simple data access and subsetting from a remote THREDDS server from python.

Nevertheless, there are groups at various institutions who continue to desire ldm/idd access to various data streams. Having sites external to Unidata available as IDD data relays, and as THREDDS servers is desirable so that there is not a single point of failure in the data distribution path and also simply to reduce the data serving load placed on the Unidata infrastructure and staff.

We intend to continue our involvement in providing large data resources to those interested. We therefore propose to upgrade our existing IDD ingest and relay hardware and THREDDS server as a resource to the Unidata community. These servers would make available via ldm/IDD and THREDDS all data available on the existing Unidata IDD data feeds, including the GOES-16 and GOES-17 data sets, as well as additional data sets that will be available in the future.

Project Description

A1. Details of equipment requested

We propose to purchase an IDD relay cluster consisting of 2 servers to replace our existing circa 2013 IDD relay machine. We also propose to purchase a community THREDDS/ADDE server with sufficient RAM and disk storage to hold at least a week's worth of recent IDD data, making GOES-16 and GOES-17 data as well as existing model and other data available.

A2. Justification for equipment requested

The servers for the IDD relay cluster will have a substantial amount of RAM, to enable an ldm product queue capable of holding at least one hours worth of all LDM data feeds. The machines will also have open memory slots available for future RAM increases if necessary. One server will act as a combination director/worker node, which is where the ldm feed requests come in. The other server would act solely as a

worker node, with feed requests balanced out among the workers by the director node. This configuration will allow us to spread the data load over several machines and network connections, to reduce bandwidth congestion and enhance availability. Additional worker nodes can be added in the future if there is demand.

Our existing idd ingest and relay machine, purchased in 2013 and upgraded in RAM in 2015, consists of a dual processor AMD Opteron 6128 (16 total cores) with 96 Gb of RAM. We currently serve various data feeds to ~40 downstream sites. This machine currently does NOT relay the GOES-16 GRB data feed, mainly due to concerns about performance. We do ingest the GOES-16 GRB feed on a side machine and relay to only a few remote sites at the current time.

The proposed ldm relay cluster machines will consist of two dual Intel E5-2630V4 processor machines with 256 Gb of RAM, expandable if necessary. A 480 Gb SSD will host the operating system and software, and a 1 Tb enterprise SATA hard drive will be available for additional small storage needs. These machines will ingest and relay the entire suite of data feeds available through the Unidata IDD.

The data server machine will have dual Intel E5-2630V4 processors, 256 Gb of RAM as well as 36 Tb of RAID6 storage (8x6Tb SAS disk drives attached to an LSI RAID card) to enable hosting at least one week's worth of all of the Unidata IDD data feeds, including recent GOES-16, GOES-17, model, radar, observation, etc. data. THREDDS and ADDE will be installed as services to make this data accessible to the department and to the Unidata community.

B. Goals of the project

The goals of this project are to continue the tradition of ingesting and relaying IDD data feeds, including GOES-16/17 data, to the Unidata community via ldm, and to also make recent data available to the community via a THREDDS/ADDE server. Geographic redundancy of servers is important, both for ldm push feeds and THREDDS/ADDE so that sites are not left without data due to a single point of failure. This project would provide geographic redundancy, as well as reduce the operational and management load placed on servers housed at Unidata in Boulder.

C. Benefits to research/education

The Department of Atmospheric and Oceanic Sciences at the University of Wisconsin-Madison (UW-AOS) has an excellent undergraduate program that graduates between 15 and 30 B.S students each year, and a graduate program that graduates approximately 15 M.S. and Ph.D. students each year. The education and training of these students involves extensive use of earth science observational data and numerical model output, along with analysis and visualization software to interrogate this data.

The research and education conducted at UW-AOS depends heavily on the availability of large observational and model data sets. The most recent new sets of observations, the next generation GOES-16 and GOES-17 data sets allow us to sample the atmosphere at a much higher spatial and temporal resolution than ever before. These, combined with ever increasing model time and space resolution, require more resources to both relay to downstream sites, and to host locally.

The upgraded IDD relay cluster and THREDDS/ADDE server will allow UW-Madison to continue to ingest and host existing and new data sets for in-house research and educational use.

D. Added value to Unidata community at large

The upgraded IDD relay server will enable UW-AOS to continue to ingest and serve data, including GOES-16 and GOES-17 data, to Unidata community sites that wish to receive data via the ldm/IDD.

Additionally, the THREDDS/ADDE server will allow us to make this data available to other Unidata sites who do not want, or are otherwise unable to afford or maintain the infrastructure and expertise necessary to run their own ldm ingest and data services. UW-AOS will be able to share the load with Unidata and other available THREDDS/ADDE servers.

Outreach programs conducted by UW-SSEC, the National Weather Service, and the American Meteorological Society and their attendees will also benefit from enhanced data availability in-house. The K-12 teachers that take advantage of these programs will have a better opportunity to become familiar with analysis and visualization tools, and incorporate these tools into their curricula. Similarly, students taking part in these programs will enter college having already been exposed to Unidata software and available data.

E. Relationship to existing computing facilities and resources, and departmental plan

At UW-AOS, we currently maintain multiple Unidata IDD ingest/relay machines that ingest the entire suite of IDD data feeds, and relay much of this data internally and to many downstream sites. The ldm relay cluster will enhance our ability to provide this service both internally and to the Unidata community.

Our existing real-time and internal archive data server is inadequate to host the entire GOES-16-17 data sets in addition to the observational and model data that is currently hosted. The new RAMADDA/THREDDS server would act as a new departmental data server as well as a resource for the Unidata community.

Budget

The two (2) ldm relay servers cost $5569.06 each (total of $11138.12 for the two)

The THREDDS/ADDE server costs $8305.96

The total cost for all of the proposed hardware is $19444.08.

Setup and maintenance of the proposed hardware will be done by the UW-AOS systems administrator as part of his normal job duties.

Project Milestones

The proposed equipment will be ordered as soon as possible prior to the summer of 2018. The servers will be installed and configured in early summer 2018, and should be ready for use by UW-AOS courses and the greater Unidata community before the fall semester, 2018.

ACECOMPUTERS

ISO 9001

Quotation

Ace Computers

575 Lively Blvd
Elk Grove Village, IL 60007
(847)952-6900, FAX: (847)952-6901
http://www.acecomputers.com

NASA SEWP V Contract #NNG15SC01B
UW Madison Contract 16-5882
GSA Schedule GS35F-0400T, Expires 04/23/22
CAGE Code 1PWX7

johns@acecomputers.com

intel Technology Provider Platinum 2018
HPC Data Center Specialist

Microsoft PLATINUM OEM

NVIDIA PARTNERFORCE TESLA PREFERRED PROVIDER

AMD Fusion Partner Select

Microsoft Partner Gold OEM Hardware
Authorized Academic
Authorized Open Licensing
Authorized Storage Server

Date:	3/13/2018
Quote By:	John Samborski, PE
Page	1 of 1
REV:	02

To: UW-Madison Dept of Atmospheric Sciences
Pete Pokrandt

Server Requested

Qty	Item	Description	Unit Price	Total Price
1	**AC-PW29925RE1CR12HE526321284TA8** 	**Ace Powerworks Server 29925E1CRBH with Dual Intel Xeon E5-2630V4 Processors** 2U Server with 12 Hot pluggable drive bays and redundant 920W Power supplies Dual Intel Xeon E5-2630V4 Processors - 10 Cores/20 Threads each, 2.2 GHz/3.1 GHz 256GB DDR4 ECC Registered memory, 8 x 32GB DDR4 DIMMS, 2400 MHz ECC Registered (8 open DIMM slots for future) LSI 3108 RAID Controller with 2GB Cache 8 x 6TB SAS 7200 RPM Hard Drives, Set as RAID 6 Dual 10GB Ethernet ports using Intel X540 plus IPMI 2.0 / KVM over LAN / Media over LAN 3 year warranty on parts and labor	$ 8,305.96	$ 8,305.96
2	**AC-PW19925RTDWE52632128250S2TA** 	**Ace Powerworks Server 19925RTDW with Dual Intel Xeon E5-2630V4 Processors** 1U Server with 4 Hot pluggable drive bays and 600W Platinum Power Supply Dual Intel Xeon E5-2630V4 Processors - 10 Cores/20 Threads each, 2.2 GHz/3.1 GHz 256GB DDR4 ECC Registered memory, 8 x 32GB DDR4 DIMMS, 2400 MHz ECC Registered (8 open DIMM slots for future) 480GB SSD plus 1TB Enterprise SATA Hard Drive Dual GB Ethernet ports using Intel i350 plus IPMI 2.0 / KVM over LAN / Media over LAN 3 year warranty on parts and labor	$ 5,569.06	$ 11,138.12

COMMENTS: FOB Destination

John Samborski, PE

Sub-total:	$ 19,444.08
Shipping:	$ -
Parts Total:	$ 19,444.08
Tax:	$ -
Page Total	$ 19,444.08
Terms	NET 30 Days