

NSF Unidata Community Equipment Awards Cover Sheet

Upgrading the community THREDDS data server and LDM/IDD relay infrastructure at UW-Madison AOS

Date: April 8, 2024

Principal Investigator Name: Ankur Desai

Title: Department Chair

Department: Department of Atmospheric Sciences – University of Wisconsin – Madison

Telephone Number: (608) 265-9201

Street Address: 1225 W. Dayton St, Madison, WI 53706

Email address: desai@aos.wisc.edu

Signature of PI: _____



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 6301

Madison, WI 53715-1218

Telephone Number: (608) 262-3822

FAX Number: (608) 262-5111

Email address: preaward@rsp.wisc.edu

Signature of University Official: _____

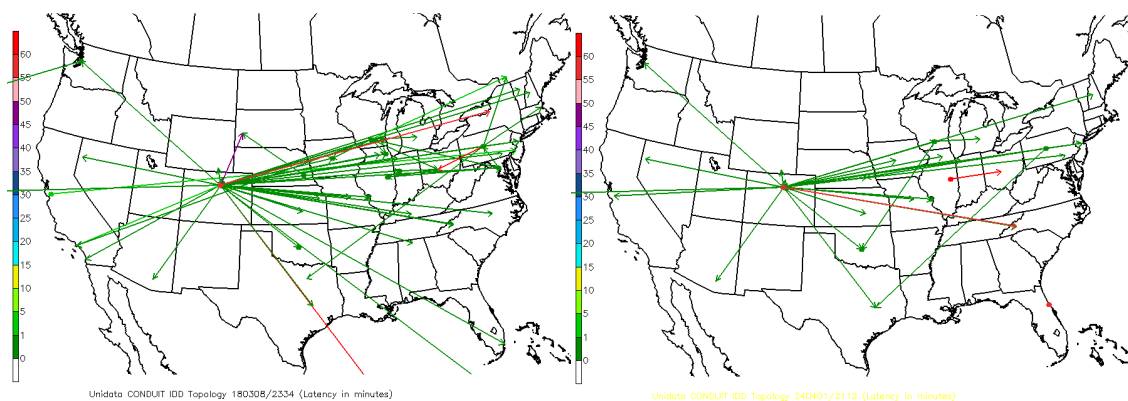


Project Summary

Since the early days of the NSF Unidata Internet Data Distribution project (IDD), the UW-Madison Department of Atmospheric and Oceanic Sciences (UW-AOS) has served as a top-level node for IDD data distribution, and has maintained a local data server used in-house with GEMPAK, the IDV, etc. Data saved at UW-Madison has frequently been offered to other participating NSF Unidata sites to fill in data gaps or help with research or case study data needs. Since 2018, thanks to a successful NSF Unidata Equipment grant, UW-AOS has also hosted a THREDDS server (thredds.aos.wisc.edu) with content mostly mirroring that of thredds.ucar.edu. This THREDDS server provides an additional repository to the community of most of the data that flows over the IDD. That NSF Unidata Equipment grant also funded a small IDD relay cluster, which has provided data via the IDD to a variety of NSF Unidata institutions.

Continuing the trend of 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, and GFS, among others, are available at increasingly high time and spatial resolutions. Similarly, the higher time and space resolution, as well as the 16 frequency bands available from the GOES-R series of satellites have generated 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 NSF Unidata directly, and not relaying to other sites. Figure 1 shows a stark example of this using CONDUIT data from March 8, 2018 and an update from April 1, 2024. The vast majority of sites are feeding direct from NSF Unidata's servers in Boulder.



(Figure 1 - Left - CONDUIT topology for March 8, 2018, Right – same for April 1, 2024)

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 are utilized. These technologies allow for large data sets to be housed on remote servers and accessed and subsetted 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 code.

Nevertheless, there are groups at various institutions who continue to desire LDM/IDD access to various data streams. Having sites external to NSF 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 NSF Unidata infrastructure and staff. The recent announcement that the NSF Unidata ldm server at the Space Science and Engineering Center (SSEC) at UW-Madison will be permanently decommissioned means there will now be one less alternate site available for IDD nodes to receive data from.

We at UW-AOS intend to continue our involvement in providing large data resources to those interested. We therefore propose to upgrade our existing 6 year old IDD ingest and relay hardware and THREDDS servers and continue to provide them as resources to the broader NSF Unidata community. These servers will continue to make available via ldm/IDD and THREDDS all data available on the existing Unidata IDD data feeds, including the GOES-R series data sets, as well as additional data sets that will likely 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 2018 IDD relay machines. We also propose to purchase a community THREDDS server with sufficient RAM and disk storage to hold at least a week's worth of recent IDD data, making GOES-R series data as well as existing model, RADAR 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. Both servers will act solely as worker nodes, with feed requests balanced out among the workers by an existing director node. This configuration will allow us to spread the data load over both 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 worker node machines, purchased in 2018 consist of a dual processor Intel Xeon E5-2630 v4 (20 total cores) with 256 Gb of RAM. We currently serve various data feeds to ~40 downstream sites.

The proposed ldm relay cluster machines will consist of two single intel Xeon Silver 4516Y processor (24 cores/48 threads) machines with 512 Gb of RAM, expandable if necessary. A 960 Gb SSD will host the operating system and software. These machines will ingest and relay the entire suite of data feeds available through the NSF Unidata IDD.

The data server machine will have a single AMD EPYC 7313P processor (16 cores/32 threads), 512 Gb of RAM as well as 108 Tb of RAID6 storage (8x18Tb SATA disk drives attached to a Broadcom/Avago 3108-8i RAID controller with 1 Gb cache) to enable hosting at least one week's worth of all of the NSF Unidata IDD data feeds, including recent GOES-R series satellite, model, NEXRAD, observation, etc. data. THREDDS will be installed as services to make this data accessible to the department and to the NSF Unidata community.

B. Goals of the project

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

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, a one year professional M.S. program that graduates between 8 and 15 students per year, and a research 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 GOES-R series satellite 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 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 NSF Unidata community at large

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

Additionally, the upgraded THREDDS server will continue to allow us to make this data available to other NSF 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 NSF Unidata and other available THREDDS 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 NSF Unidata software and available data.

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

At UW-AOS, we currently maintain multiple NSF 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 upgraded ldm relay cluster will enhance our ability to continue to provide this service both internally and to the NSF Unidata community.

Additionally, the upgraded THREDDS server would continue to act as a departmental data server as well as a resource for the Unidata community.

Budget

Equipment: \$20,000

Two (2) new ldm relay servers as proposed - \$6,236.90 each (total of \$12,473.81 for the two)

THREDDS data server as proposed - \$7,994.71

The total cost for all of the proposed hardware is \$20,468.52 (quote attached).

Indirect costs / Facilities and Administration (F&A) costs - \$0.00

The current University of Wisconsin Indirect Cost Rate (F&A) is 55.5%, which will be in effect through June 30, 2026 per the negotiated rate agreement updated November 21, 2023. Permanent equipment over \$5000 are excluded from the F&A cost base.

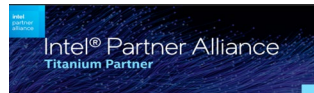
UW-AOS will supplement awarded funds to cover the \$468.52 exceeding the NSF Unidata Equipment grant limit of \$20,000. Setup and maintenance of the proposed hardware will be done by the UW-AOS systems administrator as part of his normal job duties.

Project Milestones

The proposed equipment will be ordered as soon as possible prior to the summer of 2024. The servers will be installed and configured in early summer 2024, and should be ready for use by UW-AOS courses and the greater NSF Unidata community before the fall semester, 2024.



Ace Computers



**HPC Data Center Specialist
Education Specialist**

340 Howard Avenue
Des Plaines, IL 60018
(847)952-6900, FAX: (847)952-6901
<http://www.acecomputers.com>
NASA SEWP V Contract #NNG15SC01B
UW Contract #UWMSN-2023-1134-IT Hard
GSA Schedule GS35F-0400T, Expires 04/23/27
NASPO Valuepoint Contract 2023-2028
CAGE Code 1PWX7
johns@acecomputers.com



Date:
Quote By:
Page
Rev

4/4/2024
John Samborski
1 of 1
02



To: Pete Pokrandt
Dept of Atmospheric and Oceanic Sciences
University of Wisconsin - Madison

Data Ingest/Relay and Storage Servers

Qty	Item	Description	Unit Price	Total Price
2	AC-PWKS1E6330S4516200480S	<p><u>Ace Powerworks 1E6330S Server with Intel Xeon Silver 4516Y Processor</u> 1U Edge Type Server with 800W Titanium Rated Redundant Power Supplies 1 x Intel® Xeon® Silver 4516Y with 24 Cores/48 Threads, 2.2 GHz Base/3.7 GHz Turbo Speed, 45 MB Cache 512 GB DDR5 RAM, 4800 MHz, ECC Registered, 64GB x 8 (8 DIMM slots open) 960GB M.2 NVMe Enterprise Class Drive Installed (Micron 7450 Pro - Read Intensive 5000 MB/S Read/1400 MB/S Write) 2 x 2.5" Gen4 NVMe/SATA Hot Swap Bays - Front Mounted - Open for expansion 2 x PCI-E Gen 5 x 16 Full Height, Half Length Slots 2 x OCP 3.0 Gen 5 x 16 Slots Integrated Video in ASPEED AST2600 2D Video Adapter/PCI-E Interface Dual 1GB Base-T LAN w/ Intel i210 and dedicated management port with KVM over LAN 3 Year Warranty on parts and Labor</p>	\$ 6,236.90	\$ 12,473.81
1	AC-PSS272ZR30160	<p><u>Ace PowerStor Storage Server with 100TB RAID 6 Storage</u> 2U Rackmount Storage Server platform with Dual Redundant 800W Platinum Redundant Power Supplies 1 x AMD® EPYC® 7313P, 16 Cores/32 Threads, 3.0 GHz/3.7 GHz Turbo, 128MB Cache 512GB DDR4 ECC Registered memory, 3200 MHz (8 x 64GB DIMMS), 8 Open slots remain Dual Gigabit LAN w/ Intel® i350 and dedicated management port with KVM over LAN Broadcom/Avago 3108-8i RAID Controller with 1GB Cache 8 x 18TB Ultrastar DC Enterprise SATA Hard Drives, 7200 RPM (108TB in RAID 6) 960GB M.2 NVMe Enterprise Class Drive Installed (Micron 7450 Pro - Read Intensive 5000 MB/S Read/1400 MB/S Write) 3 Year Warranty on parts and Labor</p>	\$ 7,994.71	\$ 7,994.71

COMMENTS: FOB Destination

John Samborski, PE

Sub-total:	\$	20,468.52
Shipping:	\$	-
Parts Total:	\$	20,468.52
Tax:	\$	-
Page Total	\$	20,468.52
Terms		Net 30 Days