A Proposal to the

2016 Unidata Community Equipment Awards

for Support of

A Pilot Project for Cloud-based NEXRAD Data Processing, Analysis, and Visualization for Flood Forecasting and Water Resources Management

3/22/2016

Principal Investigator

Dr. Bong Chul Seo
Assistant Research Scientist
IIHR—Hydroscience & Engineering
The University of Iowa
100 C. Maxwell Stanley Hydraulics Lab.
Iowa City, IA 52242
Phone: 319-384-2041
Fax: 319-335-5238
Email: bongchul-seo@uiowa.edu

Institution Official

Dr. Daniel Reed
Vice President for Research
The University of Iowa
2 Gilmore Hall
Iowa City, IA 52242
Phone: 319-335-2123
Fax: 319-335-2130
Email: era@uiowa.edu

Paul Below
Acting for Daniel Reed

Dr. Daniel Reed
VP for Research and Economic Development

Dr. Bong Chul Seo
Principal Investigator

Project Summary

This pilot project will build a prototype system that processes, analyzes, and visualizes NEXRAD radar data on the cloud. The project will build on the past development of Hydro-NEXRAD software system that enabled users to access vast amounts of NEXRAD radar data in support of a wide range of research. The system processes basic radar data (Level II) and delivers radar-rainfall products based on the user's custom selection of features such as space and time domain, river basin, rainfall product space and time resolution, and rainfall estimation algorithms. This was a collaborative research project among academic institutions (University of Iowa and Princeton University) and research centers (UCAR's Unidata and NOAA's NCDC), with support from the National Science Foundation. The project was conducted in two phases to handle both historic archive and real-time (e.g., Hydro-NEXRAD-2) data. The prototype system contained data from some 40 radar locations over the United States, and the software included several innovations such as very efficient custom data format, relational database of comprehensive metadata, map-based interface, and custom algorithms designed with hydrologic user in mind. At the peak the system had over 200 users representing a wide range of organizations. Unfortunately, the Hydro-NEXRAD system does not currently exist due to a fatal system crash and lack of funding for system fix, maintenance, and support (e.g., incessant archive data ingestion and rapid growth of data storage).

Recently, the real-time and full historical archive of NEXRAD Level II data, covering the entire United States from 1991 to present, became available on Amazon cloud S3. This new opportunity can eliminate prior challenges faced by Hydro-NEXRAD data acquisition and processing: (1) temporal and spatial limitation arising from the limited data storage; (2) archive (past) data ingestion and format conversion; and (3) separate data processing flow for the past and real-time Level II data. Therefore, the purpose of this pilot project is to rebuild Hydro-NEXRAD capabilities on the cloud, thus eliminating prior constraints (limited space and time domain), to examine the feasibility and performance of implementing MPI-based processing and computations to the cloud, and to assess response times from an interactive web-based system. This meets the Unidata Program emphases on cloud-based storage, data analysis, and visualization, as well as the use of Python for data proximate analysis and visualization in the cloud. The cloud resources requested for this pilot project will also be used for hydrology and water resources classes to help students understand the role of precipitation and its importance in hydrology and the environmental sciences.

Project Description

Precipitation is a main driving factor of numerous hydrologic processes, and reliable rainfall information with proper space and time scale is a key element for managing floods, water supply, water quality, etc. For several decades, weather radar has been used for weather monitoring and it has provided quantitative rainfall information with fairly high space and time resolutions over large areas. These quantitative rainfall or precipitation estimates (QPE) are used to drive models that describe hydrologic (e.g., rainfall-runoff) and environmental (e.g., agricultural non-point source pollution) processes. Since its first deployment in the early 1990s, the U.S. Weather Surveillance Radar-1988 Doppler (WSR-88D) radars, known as NEXRAD, have enhanced observational capabilities in the face of increasing demands for improved resolution and greater accuracy in QPE. This radar QPE has become a primary dataset that drives a model and helps represent complexities associated with the space and time variability of hydrologic processes (e.g., distributed modeling).

The official QPE product of the U.S. National Weather Service is served with a fixed space and time resolution (e.g., 4 km and 1-h), and thus it is hard to meet requirements for a range of various hydrologic applications that require higher resolution data. To promote and facilitate the wider use of radar data in hydrologic research, a team of researchers developed a browser-based software system called Hydro-NEXRAD (Krajewski et al., 2011). The system allowed users to generate custom rainfall product using NEXRAD radar Level II data. The Hydro-NEXRAD system was a large-scale prototype system that managed and organized data from a select group (about 40 out of 160) of WSR-88D radars with a limited data period (ending in mid-2008). Radar- and basin-centric metadata (e.g., daily rainfall statistics) were also provided to help the user's selection (Kruger et al., 2011). Hydro-NEXRAD-2 (Krajewski et al., 2013) was then implemented to add real-time capability based on the Unidata Local Data Manager (LDM) and Internet Data Distribution (IDD) technique, removing the spatial limitation.

Since both the real-time and full historic archive NEXRAD Level II data over the entire United States have recently became available through Amazon S3, the Level II data can be directly acquired from S3 without the data ingestion and format change that were previously required for the Hydro-NEXRAD processing. This eliminates significant prior challenges regarding system data storage and Level II ingestion, as well as space and time domain restrictions. This pilot study will begin by implementing Hydro-NEXRAD capabilities on the cloud using the new polarimetric features (e.g., Seo et al., 2015), as well as the existing modules (e.g., Seo et al., 2011) and Python scripts. We will limit the spatial domain to Iowa to enhance data processing and computation efficiency in the pilot project. We will then test the reliability and feasibility of cloud computation and processing, followed by an assessment of response times from an interactive web-based system. Upon user request, the cloud-based system will first search Level II data using the THREDDS data server installed on S3 and then process according to workflow and algorithm specifications. User will be able to analyze and visualize the requested data and products. Real-time data acquisition through LDM and THREDDS will also be tested and evaluated. The pilot system will be open to research communities, and requested products can be delivered with the option of standardized NetCDF format through URL, FTP, or LDM.

Regarding visualization of the data, we have developed a unique web-based browser of Level II data that we have briefly described in Demir et al. (2015). The browser allows inspecting radar volume scan data in several way avoiding any interpolation. It works much faster than the Unidata's IDV (Murray et al., 2003). For visualizing the rainfall products, we will adopt the visualization we have developed and deployed at the Iowa Flood Center in the Iowa Flood Information System (see Demir and Krajewski 2013; Demir et al., 2015).

This pilot project will also use the cloud for diverse hydrology and water resources classes (e.g., Introduction to Earth Science, Hydrology, Hydrometeorology, Remote Sensing, Groundwater, Water Resources Engineering, and Watershed Hydrology and Ecohydraulics) for undergraduate and graduate students in courses in the Department of Civil and Environmental Engineering at the University of Iowa. The project will help students understand the role of precipitation and its importance in hydrology and environmental sciences. This will be a great opportunity for CEE students to access radar observation data and explore the potential of using rainfall products at high space and time resolution using Unidata tools. In particular, the proposed cloud system will enable students to obtain hands-on experience manipulating and interpreting real-time data with their own knowledge and diagnostics. IIHR-Hydroscience & Engineering is a research institute within the College of Engineering of the University of Iowa, and many graduate and undergraduate students currently participate in rainfall-runoff and water quality modeling. This effort to build a cyberinfrastructure system that integrates cloud capability to provide quantitative rainfall information for hydrologic modeling will contribute to students' education.

A number of research papers have cited Hydro-NEXRAD products (e.g., Volkmann et al., 2010; Smith et al., 2011; Wright et al., 2013). We provided the generated radar-rainfall products for various research purposes: radar-rainfall uncertainty characterization and uncertainty propagation through a hydrologic model for the University of Iowa; investigation of precipitation processes over metropolitan areas for Princeton University; estimation of daily soil erosion for Iowa State University; and spatial data representation in databases for Drexel University. Moreover, the Iowa Flood Information System (http://ifis.iowafloodcenter.org) uses the real-time rainfall product feeds, which are also provided for the University of Illinois at Urbana-Champaign and the University of North Carolina. The outcomes (e.g., tools and data) of this pilot project will support development of major research grants (e.g., NSF) that further extend the pilot spatial domain (e.g., Iowa) to the national scale based on the cloud-based information infrastructure. This work requires collaboration between the University of Iowa and the Unidata teams (e.g., THREDDS and Python).

Proposed Equipment

The real-time radar Level II data currently come via the Local Data Manager (LDM). The Level II data (approximately 10GB/day) received from several radars overlooking Iowa are converted by our codes into statewide rainfall maps (3GB/day) updated every five minutes. We obtain the archive data through the NCEI upon request; one-year data from the radars are usually over 3TB. The rainfall processing system currently runs scientific modules and computational algorithms on a HPC cluster of the University of Iowa.

We propose to use the cloud systems (on the Amazon Cloud; see configuration in the budget) for the set-up of radar data processing, LDM, and web servers. The LDM server is required for comparison of current real-time and proposed methodology (using Amazon S3) and for distribution of processed rainfall products. The proposed cloud systems will help educate students, complement our current capabilities, automate both real-time and historical data processing, and remove time and space limitations.

Personnel Expertise

PI Dr. Bong Chul Seo has been an Assistant Research Scientist at the IIHR—Hydroscience & Engineering at the University of Iowa since 2010. He is a radar hydrologist and has used his expertise of hydrometeorology and atmospheric science to process NEXRAD radar data and generate quality-rainfall estimates. He participated in the Hydro-NEXRAD ITR research project as a Ph.D. student at the University of Iowa. He has been using the Unidata IDD/LDM technology to acquire radar Level II data in real-time and is also familiar with NetCDF and associated Unidata tools. He attended the Unidata training workshop (LDM in 2012 and AWIPS II in 2014).

Co-PI Dr. Ibrahim Demir is a Research Faculty Engineer at the Iowa Flood Center, University of Iowa. He currently serves at CUAHSI Informatics Committee, NSF EarthCube Technology and Architecture Committee, Unidata User Committee, and Joint Committee on Hydroinformatics (IWA/IAHR/IAHS). Dr. Demir also active in the design and development of environmental information systems, novel scientific visualization interfaces for hydrological and meteorological data, and web-based information communication. He is the architect and lead developer of the Iowa Flood Information System, a one-stop web-based platform to provide access to flood-related data, information, maps, and visualizations.

Budget

The total requested funding for this proposal is $16,773 to provide cloud systems and data storage dedicated to cloud data access and computational processing for the proposed pilot project. We used the AWS monthly calculator to estimate the project cost; the detailed cost can be found at http://calculator.s3.amazonaws.com/index.html#r=IAD&s=EC2&key=calc-3F8666AB-AD9B-4BC3-94C5-C92B2503E759.

The cloud systems and data storage are comprised of:

	Instances	Type	Billing option	Monthly cost
Web server	1	m3.xlarge	1 year no upfront reserved	$182.50
Processing server	1	m3.2xlarge	1 year no upfront reserved	$372.30
LDM server	1	m4.large	1 year no upfront reserved	$104.39
EBS volumes		1 TB		$51.20
EBS snapshots				$97.28
AWS data transfer out				$8.91
AWS support				$100.00
				$916.58
Indirect (52.5%)				$481.20
Monthly TOTAL				$1,397.78
Total for One Year				**$16,773.36**

Project Milestones

Below is a brief timeline describing the plan of action and expected milestones for this pilot project:

- May 2016 – Award notification
- June 2016 – Create the servers on the Amazon Cloud and install required software
- August 2016 – Radar- and basin-based metadata computation and ingestion
- October 2016 – Implement Hydro-NEXRAD capabilities
- December 2016 – Set up and configure web server
- February 2017 – Start service for rainfall data request
- April 2017 – Acquire users' feedback and improve capabilities and performance

References

Demir, I. and W.F. Krajewski (2013) Towards an adaptive flood information system: Integrated data access, analysis and visualization, *Environmental Modeling & Software*, 50, 77–84.

Demir, I., H. Conover, Y. He, M. McHeniry, W.F. Krajewski, B.-C. Seo, R. Goska and W. Petersen (2015) Data enabled field experiment planning, management and research using cyberinfrastucture, *Journal of Hydrometeorology*, 16(3), 1155-1170.

Krajewski, W.F., A. Kruger, J.A. Smith, R. Lawrence, C. Gunyon, R. Goska, B.-C. Seo, P. Domaszczynski, M.L. Baeck, M.K. Ramamurthy, J. Weber, A.A. Bradley, S.A. DelGreco, and M. Steiner (2011) Towards better utilization of NEXRAD data in hydrology: An overview of Hydro-NEXRAD, *Journal of Hydroinformatics*, 13(2), 255-266.

Krajewski, W.F., A. Kruger, S. Singh, B.-C. Seo, and J.A. Smith (2013) Hydro-NEXRAD-2: Real-time access to customized radar-rainfall for hydrologic applications, *Journal of Hydroinformatics*, 15(2), 580-590.

Kruger, A., W.F. Krajewski, P. Domaszczynski, and J.A. Smith (2011) Hydro-NEXRAD: Metadata computation and use, *Journal of Hydroinformatics*, 13 (2), 267-276.

Murray, D., J. McWhirter, S. Wier, and S. Emmerson (2003) The integrated data viewer: A Web-enabled application for scientific analysis and visualization, Preprints, *19th International Conference on IIPS for Meteorology, Oceanography and Hydrology*, American Meteorological Society, Long Beach, CA.

Seo, B.-C., W.F. Krajewski, A. Kruger, P. Domaszczynski, J.A. Smith, and M. Steiner (2011) Radar-rainfall estimation algorithms of Hydro-NEXRAD, *Journal of Hydroinformatics*, 13(2), 277-291.

Seo, B.-C., W.F. Krajewski, and K.V. Mishra (2015) Using the new dual-polarimetric capability of WSR-88D to eliminate anomalous propagation and wind turbine effects in radar-rainfall, *Atmospheric Research*, 153, 296-309.

Smith, J.A., G. Villarini, and M.L. Baeck (2011) Mixture distributions and the hydroclimatology of extreme rainfall and flooding in the Eastern United States, *Journal of Hydrometeorology*, 12(2), 294-209.

Volkmann, T.H.M., S.W. Lyon, H.V. Gupta, and P.A. Troch (2010) Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain, *Water Resources Research*, 46, W11554, doi:10.1029/2010WR009145.

Wright, D.B., J.A. Smith, G. Villarini, M.L. Baeck (2013) Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition, *Journal of Hydrology*, 488, 150-165.