

# **A Cloud Computing Service for Remote Sensing Data Reprojection and Science Reduction**

**Jie Li<sup>1</sup>, Youngryel Ryu<sup>2</sup>, Keith Jackson<sup>3</sup>, Deb Agarwal<sup>3</sup>, Marty Humphrey<sup>1</sup>, Catharine van Ingen<sup>4</sup>**

<sup>1</sup>Department of Computer Science, University of Virginia, Charlottesville, VA; <sup>2</sup>Department of Environmental Science, Policy and Management, University of California, Berkeley, CA; <sup>3</sup>Lawrence Berkeley National Lab; <sup>4</sup>Microsoft Research, eScience Group and Berkeley Water Center

The combination of internet availability of remote sensing data and emerging cloud computing technologies are enabling new environmental science research. Data products from satellites such as MODIS or GRACE augment ground based measurements from networks such as Fluxnet or WATERS. Cloud computing technologies such as Microsoft's Azure or Amazon's EC2 offer the necessary computational, storage, and networking resources to process these large data. However, a significant data entry barrier remains due to the difficulty in gathering and subsequently harmonizing imagery data from heterogeneous sources. To lower this barrier, we have prototyped a general image processing pipeline in the cloud using Azure. Our pipeline automates data upload from known repositories such as the LPDAAC, image reprojection to common space and time representation, hosting an optional reduction algorithm provided by the scientist, and final result delivery to the scientist desktop. We have used our Azure pipeline to compute a single science variable dealing with evapotranspiration from 9 different MODIS products across the US over 10 years. This computation required uploading and reprojecting approximately 4 TB of imagery consuming 70K cpu hours to deliver approximately 50GB resulting images to the scientist desktop. We believe our pipeline can be generalized to various remote sensing data products analyses other than the MODIS.