Investigation into Cloud Computing for More Robust Automated Bulk Image Geoprocessing

Richard B Brown, CSC; James C Smoot, CSC; Dr. Lauren Underwood, CSC

Geospatial resource assessments frequently require timely geospatial data processing that involves large multivariate remote sensing data sets. In particular, for disasters, response requires rapid access to large data volumes, substantial storage space and high performance processing capability. The processing and distribution of this data into usable information products requires a processing pipeline that can efficiently manage the required storage, computing utilities, and data handling requirements. In recent years, with the availability of cloud computing technology, cloud processing platforms have made available a powerful new computing infrastructure resource that can meet this need. To assess the utility of this resource, this project investigates cloud computing platforms for bulk, automated geoprocessing capabilities with respect to data handling and application development requirements.

This presentation is of work being conducted by Applied Sciences Program Office at NASA-Stennis Space Center. A prototypical set of image manipulation and transformation processes that incorporate sample Unmanned Airborne System data were developed to create value-added products and tested for implementation on the ”cloud”. This project outlines the steps involved in creating and testing of open source software developed process code on a local prototype platform, and then transitioning this code with associated environment requirements into an analogous, but memory and processor enhanced cloud platform. A data processing cloud was used to store both standard digital camera panchromatic and multi-band image data, which were subsequently subjected to standard image processing functions such as NDVI (Normalized Difference Vegetation Index), NDMI (Normalized Difference Moisture Index), band stacking, reprojection, and other similar type data processes. Cloud infrastructure service providers were evaluated by taking these locally tested processing functions, and then applying them to a given cloud-enabled infrastructure to assesses and compare environment setup options and enabled technologies. This project reviews findings that were observed when cloud platforms were evaluated for bulk geoprocessing capabilities based on data handling and application development requirements.

Investigation into Cloud Computing for More Robust Automated Bulk Image Geoprocessing[1]

Richard B. Brown, CSC, Senior Scientist
James C. Smoot, CSC, Scientist Principle Leader
Lauren W. Underwood, CSC, Scientist Leader
C. Duane Armstrong, NASA, Chief, ASTPO
NASA Applied Science & Technology Project Office (ASTPO), Stennis Space Center,
Stennis MS 39529-0001
richard.b.brown@nasa.gov
james.c.smoot@nasa.gov
lauren.w.underwood@nasa.gov
curtis.d.armstrong@nasa.gov

ABSTRACT

Geospatial resource assessments frequently require timely geospatial data processing that involves large multivariate remote sensing data sets. For disasters, response requires rapid access to large data volumes, substantial storage space and high performance processing capability. Data processing and distribution requires a processing pipeline that can efficiently manage the necessary storage, computing utilities, and data handling so that products are usable. Recent expansion of ready availability cloud computing technology capabilities that can accept a greater multitude of operating systems, application programming interfaces (APIs), and web based processing tools has enabled a powerful new computing infrastructure resource. The utility of cloud computing platforms for automated geoprocessing capabilities (data handling and application development requirement), were investigated in this study.

A prototypical set of image manipulation and transformation processes using sample Unmanned Airborne System (UAS) data acquired from NASA Ames Research Center were developed to create value-added products and tested for implementation on the "cloud." Initial steps involved creating and testing open source software on a local prototype platform, and then transitioning this code, with associated environment requirements, into an analogous, but memory and processor enhanced cloud platform. A NASA Nebula data processing cloud instance was then used to store both standard digital camera panchromatic and multi-band image data, which were subsequently subjected to standard image index processing functions such as NDVI (Normalized Difference Vegetation Index), and mosaicing. Findings observed on Nebula were evaluated for bulk geoprocessing capabilities based on data handling, application development requirements, and processing speed.

Key words: Cloud computing, Geoprocessing, Geographic Information Systems (GIS), Image Processing, Nebula, Unmanned Airborne System (UAS), Remote Sensing

INTRODUCTION

A prototypical set of image manipulation and transformation processes that incorporate sample UAS data to create value-added products and test for implementation on the "cloud" was developed. A data processing cloud instance was used to store standard digital panchromatic and multi-band image data, which was subsequently subjected to standard image processing functions. A processing flow that describes the real-time airborne sensor monitoring and data-flow/processing validation techniques utilizing cloud computing capabilities is shown in Figure 1.

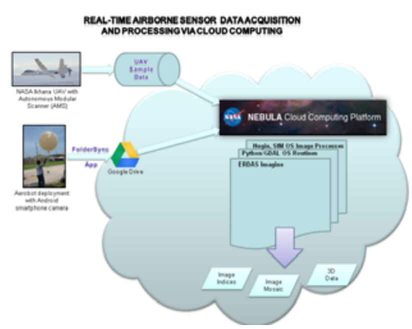

Figure 1 Data flow for cloud computing capabilities

[1] This manuscript is a joint work of employees of the National Aeronautics and Space Administration and employees of CSC under Contract No. [NNS10AA35C] with the National Aeronautics and Space Administration. The United States Government may prepare derivative works, publish or reproduce this manuscript, and allow others to do so, for United States Government purposes. Any publisher accepting this manuscript for publication acknowledges that the United States Government retains such a license in the published form of this manuscript.

Project Objectives

The purpose of this experiment was to demonstrate and evaluate the ability of existing cloud computing systems to acquire image data and process it in real time. UAS data were used as input for both Commercial-Off-The-Shelf (COTS) geoprocessing software packages and analogous data manipulation functions developed using open source software. The data processing was performed on local desktop systems and in a cloud computing environment. The results were used to determine feasibility and performance metrics.

Project Process Flow

1. Identify and acquire sample UAS data
    a. Simulate real-time UAS data transfer from the UAS to the cloud environment
    b. Automate data transfer from smart-sensor (Android camera) through Wi-Fi to cloud storage
    c. Alert users of incoming data
2. Proof of concept: develop and test basic data processes/tools locally (on a local server) on sample UAS data
    a. Image indices (e.g. Normalized Difference Vegetation Index, NDVI)
    b. Mosaics
    c. Reprojection
3. Cloud computing service: Nebula
4. Nebula cloud computing: develop the framework
    a. Create an Ami-Bastion host instance to serve as gateway for user access to Nebula
    b. Create a Windows 2008 Server™ instance from dashboard and specify size/storage requirements
    c. Setup permanent storage volume if needed (100GB total limit per project)
5. Nebula data storage: test various data push technologies for (Google Drive™/NASA Large File Transfer (LFT tool) moving data and software to Nebula storage volume.
6. Environment test of processes on Nebula cloud instance
    a. Install data processing tools from the local environment to the defined cloud and
    b. Perform a functional checkout of analogous COTS and open source processes on Nebula
7. Compare cloud environment computing performance with all packages/processes

1. Identify and Acquire sample UAS Data

NASA-Ames Research Center (Ambrosia, et al.) kindly provided the sample UAS data by providing a link to their Autonomous Modular Scanner (AMS) sensor data which is flown on NASA's large UAS aircraft, Ikahana. Ikahana is a General Atomics Predator B that was acquired by NASA, for use as an aeronautical research aircraft as well as to serve the Earth science community. Ikhana carries the AMS payload developed by NASA's Ames Research Center. The equipment incorporates a sophisticated imaging sensor and real-time data communications equipment. AMS is an airborne scanning spectrometer that acquires high spatial resolution imagery of the Earth's features from its vantage point on-board the Ikahana research aircraft. Data acquired by AMS is used to define, develop, and test algorithms for use in a variety of scientific programs that emphasize the use of remotely sensed data to monitor variation in environmental conditions, assess global change, and respond to natural disasters. Data collected is then downlinked to NASA Ames. Technical support regarding UAS data format discoveries for this project was also provided by NASA Ames. The AMS imagery has 12 spectral bands of information and the wavelengths are very similar to that of Landsat Thematic Mapper (TM) and Visible Infrared Imager Radiometer Suite (VIIRS); the breakdown of the band comparisons is shown in Figure 2. Using the AMS imagery as input, the cloud-enabled processing functions were tested, and data products were generated and validated.

Band	Wavelength (nm)	Simulated Band
1	420 - 450	
2	450 - 520	TM 1
3	520 - 600	TM 2
4	600 - 620	
5	630 - 690	TM 3
6	690 - 750	
7	760 - 900	TM 4
8	910 - 1050	
9	1550 - 1750	TM 5
10	2080 - 2350	TM 7
11	3600 - 3790	NPOES VIIRS M12
12	10.26 - 11.26µm	NPOES VIIRS M15

Figure 2 AMS sensor band comparison

Simulated real-time data transfer from the UAS to cloud

For this project, a prototypical set of image manipulation and transformation processes that incorporate sample UAS data to create value-added products for testing and implementation on the “cloud” was developed. A NASA cloud service (Nebula) was used to store standard digital panchromatic and multi-band image data, which in turn were subsequently subjected to standard image processing functions (Figure 3).

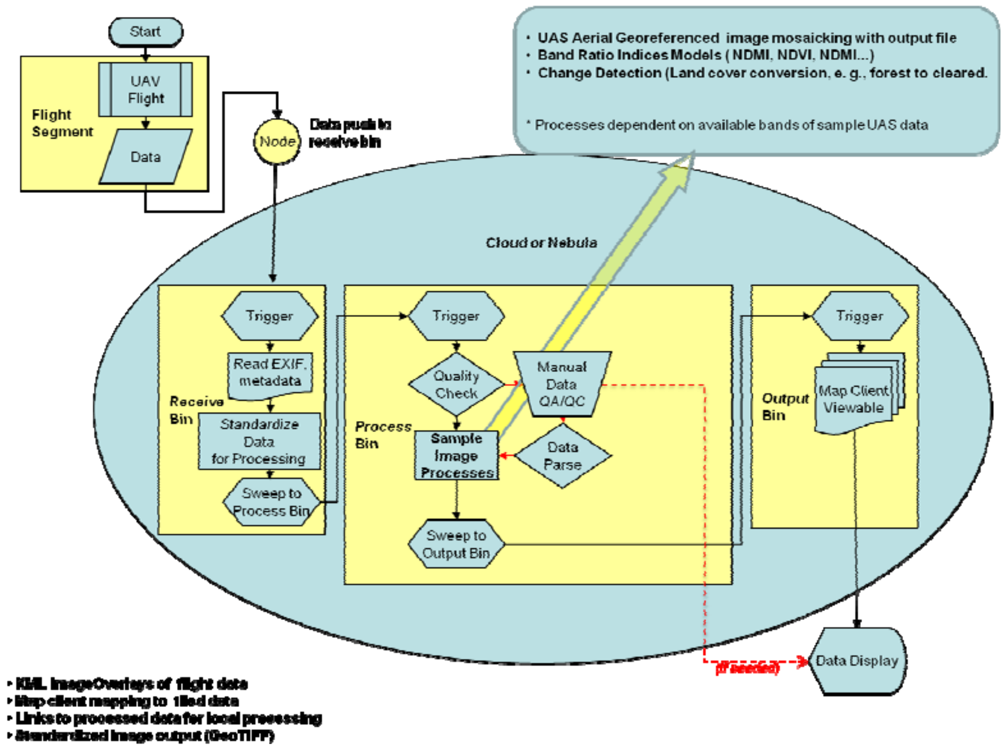

Figure 3 Process flow diagram: UAS data streaming to a cloud environment.

2. Proof of concept: Developed and test basic data processes/tools locally via open source code using UAS data and transition to a cloud environment

For the proof of concept, several “cloud-enabled” processing functions that included NDVI, NDMI (Normalized Difference Moisture Index), layer stacking, reprojection, and other similar processes were identified, developed, and tested via a local virtual environment. Python code has been successfully used to program these types of processing functions. Python is an open source programming language that is extensible for geoprocessing via, for example, the Geospatial Data Abstraction Library (GDAL)[2] and Numerical Python (NUMPY)[3] was used to process the data. The output products were checked in Opticks[4], another open source expandable remote sensing and imagery analysis software platform. For validation purposes, the same data sets were used as input in ERDAS Imagine (COTS package) to create similar type value-added products. The products created using the open source software were compared against the COTS output products, and the end results were comparable.

3. Cloud computing service: Nebula

Ultimately, Nebula[5] was the selected cloud computing service for this proof-of-concept. Nebula is an open-source cloud computing project service developed to provide an alternative to the costly construction of additional data centers required whenever NASA scientist or engineers need additional data processing

[2] http://www.gdal.org/
[3] http://numpy.scipy.org/
[4] http://opticks.org/confluence/display/opticks/Welcome+To+Opticks
[5] http://nebula.nasa.gov/

capabilities. Nebula also provides a simplified avenue for NASA scientists and researchers to share large, complex data sets with external partners and the public. NASA-Stennis Applied Science & Technology Project Office (ASTPO) was granted full access to Nebula for the implementation and testing of the processes developed for this project.

4. Build Nebula Cloud Computing Framework

NASA-Ames provided Nebula technical support relative to project requirements and for the creation of a virtual machine (VM) on the cloud (Figure 4). The VM is a software implementation of a machine (a computer) that executes programs just like a physical machine. Once the VM is established, the cloud infrastructure is set up to manage instances, images, keys, volumes and security groups (Figure 4). **Instances** are virtual servers launched from images and **images** are snapshots of running systems which can easily be deployed to run one or more instances. A requirement of a new user that is newly-assigned to a project is to create a key pair (Private Key) and save it to your local machine. **Key pairs** are Secure Shell (SSH) credentials which are ingested into images when they are launched which enables the user to connect to Nebula via a combination of SSH and remote desktop protocol (RDP). Creating a new key pair registers the public key and downloads the private key to your local machine. **Volumes** provide persistent block storage. Creating a new volume gives you a raw block device which you may format with your choice of file systems. Nebula allows the user to create and attach a volume maximum capacity of 100 GB. And finally, a **Security Group** is a named set of rules that applies to the incoming packets for the instances. For the purposes of this project an "X-large" 64 bit instance was created for testing on Nebula with 4 virtual 2.8 GHz processors and 16 GB of RAM running on a Windows 2008 Server R2 operating system.

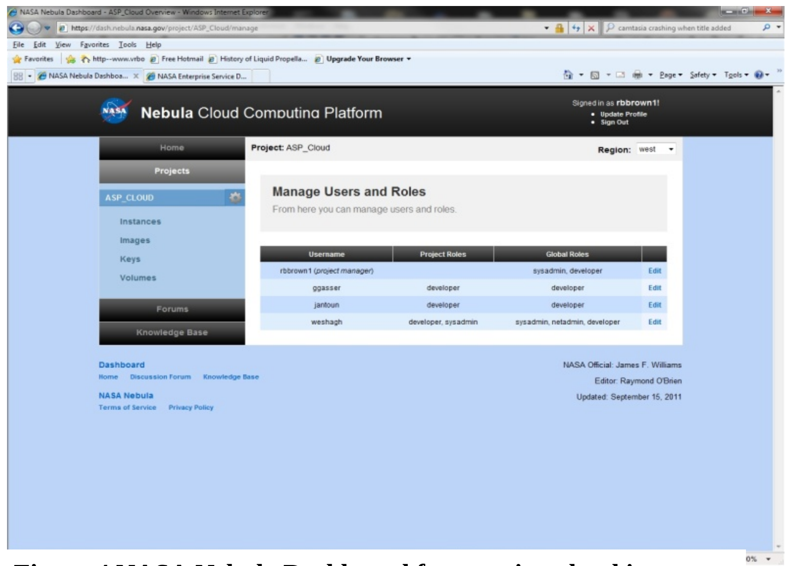

Figure 4 NASA Nebula Dashboard for creating cloud instances

5. Nebula Data Storage

Initially, when setting up a cloud VM for the first time, the user can specify the project instance storage size (small, median and large) up to 160 gigabytes (GB). However, the storage created with that instance is not permanent, and only exists as long as the instance is there. For example, if the instance is lost via a system crash, then the data stored on the instance will also be lost. Nebula does however allow the user, if needed, to maintain and preserve a 100 GB total limit, per project, of permanent volume storage; in the near future, up to 100 TB of storage will become available. This volume storage (a virtual drive) consists of standard disk storage which uses fixed-size blocks, and can be created and attached to a VM instance, as well as relocated to different instances, as needed.

Commercially available existing data storage/transfer packages were utilized to examine the functionality of transferring data and software onto the Nebula environment. Then, by using Google Drive and the NASA LFT tool, large files and software packages were successfully transferred onto the cloud. Google Drive is a file storage and synchronization service developed by Google that was released on April 24, 2012. It is a free service that enables the user to store and accesses files from anywhere e.g. on the web, on a hard drive, or on the go. Google Drive provides all users with an initial 5 GB of cloud storage. Additional storage, ranging from 25 GB up to 16 TB, can be acquired through a paid monthly subscription plan ($2.49 per month for 25 GB). In order to synchronize files on a local machine in the cloud, Google Drive software must be installed and running locally, and to date, data can be copied to the cloud storage.

An alternative option for locally moving data to the cloud was achieved by using the NASA LFT tool. LFT is a tool that was designed to send large files to other than NASA Operational Messaging and Directory (NOMAD) users or to invited individuals, outside of NASA. LFT enables the user to post up to 10GB of data that can then be downloaded to Nebula. However, data transfer rate, among both mechanism of data transfer, was not optimal. Investigations to uncover a timelier means of acquiring data from a local environment to the cloud are currently underway.

6. Test Nebula Environment

In order to test Nebula's processing performance, software packages, like ArcGIS™ and ERDAS Imagine™ (Figure 5), which utilized sample UAS and Moderate Resolutions Imaging Spectroradiometer (MODIS) data as inputs, were the primary focus for this project. ArcGIS, developed by the Environmental Systems Research Institute in Redland California, is a vector based geographic information system (GIS) that is used for designing and managing solutions through the application of geographic knowledge. ERDAS Imagine, designed by ERDAS in Atlanta, Georgia, is a remote sensing application that has raster graphic editor abilities that are used for geospatial application. The software applications mentioned above were installed on the cloud and the performance of several of their modules was tested. These identical modules were also run on a internal local computer. The results that were observed between the two were comparable. Although only preliminary testing has been conducted at this time, relative to benchmarking the performance of the software packages on the cloud VM versus an internal local environment, using Nebula, the turn-around time associated with the processing of large files, is definitely minimized.

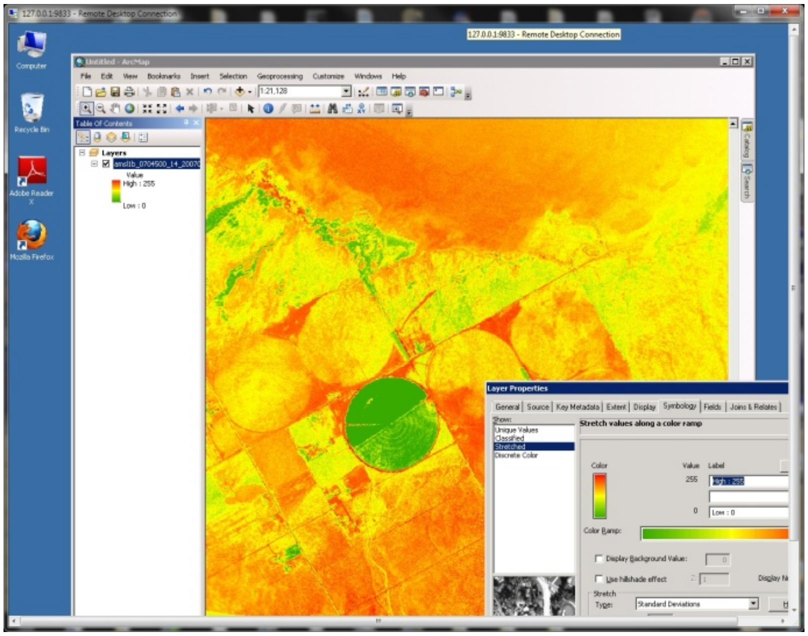

Figure 5 ERDAS Imagine NDVI product created on Nebula cloud instance

ERDAS Imagine and several open source packages were tested (including Python-GDAL scripts), and were run to produce the image indices and mosaics. The FWTools (a set of open source GIS binaries) installer package (Figure 6) was used to simultaneously load the proper co-working versions of these tools onto Nebula rather than installing them separately.

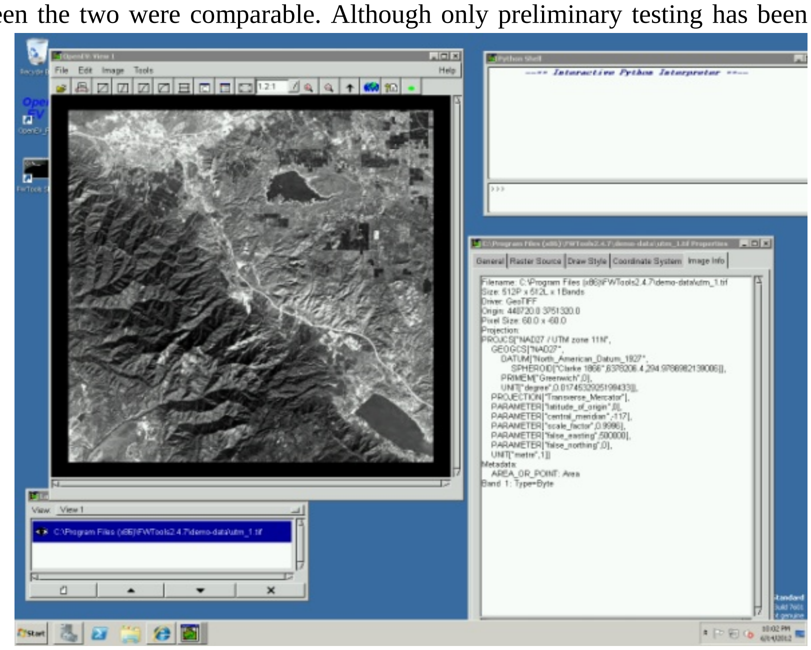

Figure 6 FWTools installed on Nebula reviewing AMS UAS data

Hugin, an open source mosaic/panorama creation tool and VisualSfM6 (Structure from Motion; see Fig 7), a point cloud extraction from overlapping imagery tool from the University of Washington, were also tested on the Android smart camera. Close range imagery was taken over local rugged barren ground test locations, and the data was successfully synced and transferred to the Google Drive cloud. Then, within 10 minutes, the data was automatically downloaded to the local storage volume on the Nebula instance. At this stage, the image processing tools could be enabled and assessed using the data that had just been delivered. Preliminary observations detected that a few of the open source tools had some graphics processor (GPU) dependencies, and they had to be switched off to use Central Processing Unit (CPU) modes only for processing, due to the fact that the Nebula cloud virtual instances do not have graphics cards installed.

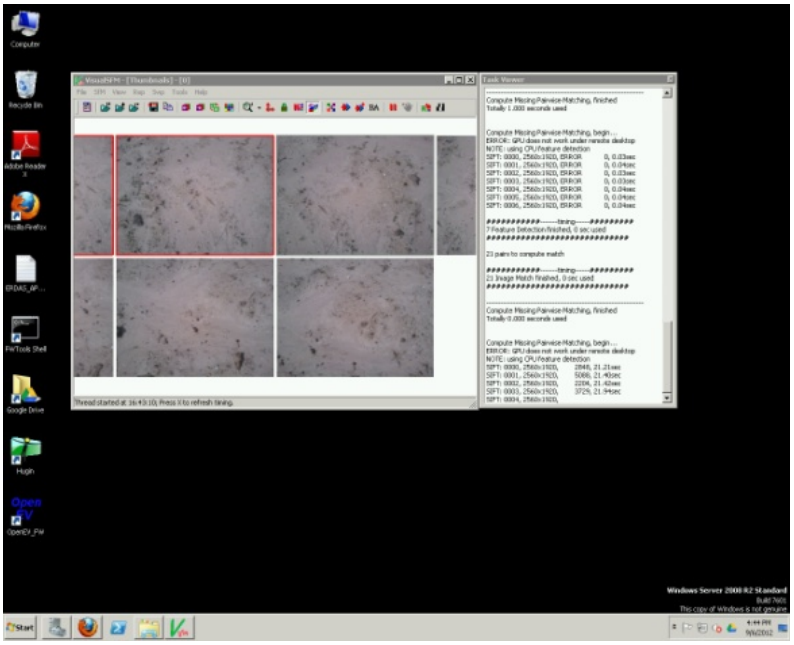

Figure 7 Open source Structure from Motion (SfM) running image feature matching routines on Android camera high res imageset

ACKNOWLEDGMENTS

We would like to thank Mr. C. Duane Armstrong, Chief, NASA's Applied Science & Technology Project Office (ASTPO), NASA Stennis Space Center, for funding and support of this project, and Mr. Vincent Ambrosia a Senior Research Scientist and principal investigator / lead scientist working with the NASA Airborne Science Program at NASA Ames Research Center for contributing UAS data to NASA ASTPO, Stennis Space Center by providing a link to AMS (Autonomous Modular Scanner) data that is downloaded from NASA's large UAS aircraft Ikahana, and stored at NASA Ames. Also much appreciation goes out to William Eshagn of NASA-Ames Research Center for his guidance and technical support provided for gainning access to the Nebula Cloud Computing Platform.

[6] http://www.cs.washington.edu/homes/ccwu/vsfm/

ABBREVIATIONS AND ACRONYMS

AMS	Autonomous Modular Scanner
API	Application Programming Interface
ASTPO	Applied Science & Technology Project Office
COTS	Commercial off The Shelf
CSC	Computer Sciences Corporation
CPU	Central Processing Unit
GB	Gigabytes
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information Systems
LFT	Large File Transfer
MODIS	Moderate Resolution Imaging Spectroradiometer
NAMS	NASA Account Management System
NDVI	Normalized Difference Vegetation Index
NOMAD	NASA Operational Messaging and Directory
NUMPY	Numerical Python
RDP	Remote Desktop Protocol
SSC	Stennis Space Center
SSH	Secure Shell
TB	Terabytes
TM	Thematic Mapper
UAS	Unmanned Airborne System
VIIRS	Visible Infrared Imager Radiometer Suite
VM	Virtual Machine

REFERENCES AND/OR SELECTED BIBLIOGRAPHY

Krabill, W.B., R.H. Thomas, C.B. Martin, R.N. Swift, and E.B. Frederick, 1995. Accuracy of airborne laser altimetry over the Greenland ice sheet, *International Journal of Remote Sensing*, *16*(7):1211-1222.

Limaye, Ashutosh S., Molthan, Andrew L., Srikishen, Jayanthi. 2010. Proposed Use of the NASA Ames Nebula Cloud Computing Platform for Numerical Weather Prediction and the Distribution of High Resolution Satellite Imagery," presentation *2010 AGU Fall Meeting*, San Francisco, CA.

Chen, A., Pham, L., Kempler, S., Theobald, M., Esfandiari, A., Campino, J., Vollmer, B., Lynnes, C. 2011. Facilitating NASA Earth Science Data Processing Using Nebula Cloud Computing, Poster at the *AGU Fall Meeting*, 5-9 December 2011, San Francisco, California.

Ambrosia, V., Hinkley, E., Brass, J., and others. 2011. The Western States UAS Fire Mission, in *Proceedings from of* the *11th Biennial Forest Service Remote Sensing Applications Conference "New Remotes Sensing Technologies for Resource Managers*." 24-28, 2006, Salt Lake City, Utah.