

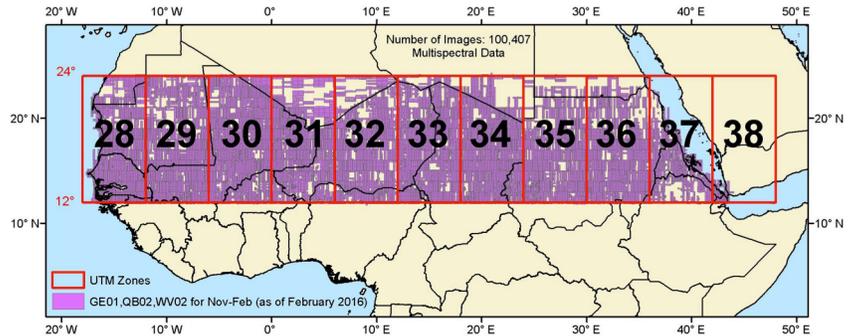


NASA Center for Climate Simulation Success Story

Counting Trees and Shrubs in the Sub-Sahara Using Cloud Computing—Part 1 July 21, 2016

At the 2014 Amazon Web Services (AWS) Public Sector Summit the Intel Corp. announced a cloud computing resource opportunity for AWS users including government agencies—bring us groundbreaking applications, and we will pay for AWS access and help to optimize the software. Sitting in the audience NASA High-End Computing Program Manager Tsengdar Lee decided to take on Intel's Head in the Clouds Challenge. His mind soon turned to Compton Tucker, a NASA Goddard Space Flight Center (GSFC) Earth scientist who pioneered satellite monitoring of vegetation more than 30 years ago. "There were tons of satellite images, and he wanted to identify certain features," Lee said. "That seemed like a - perfect cloud computing application." He called Tucker about submitting a proposal and stressed, "Think big."

Tucker and University of Minnesota colleague Paul Morin went massive. Joining Lee as principal investigators, along with the NASA Center for Climate Simulation (NCCS) and other partners, they proposed processing satellite images to count all the trees and shrubs within a coast-to-coast swath of Africa from the Sahara south to the savanna zones. "We want to perform a tree and shrub census and determine what the biological mass is," Tucker said. This biomass estimate will reveal how much carbon is stored in the region's vegetation and could be released as carbon dioxide if those plants burn or die and decompose due to natural or human causes. Moreover, it will establish a carbon baseline for later research. The science scope impressed Intel and AWS enough to offer resources. Then the real challenge began. "We had to scramble because logistically this is difficult," Tucker said.



The Sub-Sahara region of Africa covers Universal Transverse Mercator (UTM) zones 28 through 38. At nearly 10 million square kilometers the study area is larger than the continental United States. Image by Katherine Melocik, GSFC.

At base is the project's unusual scale. The Sub-Sahara stretches nearly 10 million square kilometers (km)—an area larger than the continental United States. Driving a similar distance as New York to San Francisco "imagine trying to uniquely name every tree that you see along the way," said John David, senior programmer/analyst in GSFC's Earth Sciences Division. "Now imagine trying to determine how many trees and shrubs have died or been cut down, how many are new, and when that happened over the entire record."

Getting an exact count starts with commercial satellite images acquired through the National Geospatial Agency. "We select all the images after the rainy season and before trees lose their leaves, from November to March," Tucker said. The research team started with 260,000 satellite images, totaling 200 terabytes of data. With images coming from three different satellites over multiple years, Tucker said "we had to organize the data so that we don't count it more than once."

That and other tasks fall to the NCCS private cloud—the Advanced Data Analytics Platform (ADAPT). To maximize efficiency the region gets divided into smaller and smaller pieces for parallel processing. At the top are the Sub-Sahara's 11 Universal Transverse Mercator (UTM) zones containing just over 877,000 square km apiece. UTM by UTM the researchers use ADAPT to stack all available satellite images, put them into the same map projection, select the best images, cut out overlap, and color-calibrate and resample the images to a consistent 50-centimeter resolution. These steps reduce the number of satellite images to 100,000 across all UTM zones. The team deconstructs each UTM into 100-by 100-km tiles and then 25- by 25-km sub-tiles called mosaics that go to AWS for the biomass estimate calculations.

For this demanding workload the NCCS is leveraging several innovations. They enhanced ADAPT with "fat" nodes having 256 gigabytes of random-access memory and 6 terabytes of flash memory, the latter nicely handling the myriad required input/output operations. NCCS and partner staff upgraded the network link to the AWS East facility from 1 to 10 gigabits per second. Making submissions to AWS is the data manager from Cycle Computing LLC. "It manages the links and creates multiple streams to efficiently use available network bandwidth," said Hoot Thompson, NCCS advanced technology lead.

Once at AWS East the sub-tile mosaics go onto as many as 5,000 processor cores. Team-optimized algorithms identify and count the number of trees and shrubs, measure crown area, and, from the shadows cast, determine tree height. These factors all combine into a biomass estimate for each mosaic. Additional algorithms re-assemble the sub-tile mosaics into one larger quilt-like mosaic per UTM zone. A test case with UTM zone 32—mostly the country of Niger—won the 2015 HPCwire Readers' Choice Award for the Best HPC Collaboration Between Government and Industry.

More recently the team processed the entire Sub-Sahara using panchromatic (i.e., black and white) imagery. For the AWS portion they took advantage of “spot instances” that are cheaper “if you are willing to take advantage of off-hour processing and your jobs can recover gracefully if bumped,” Thompson said. Processing 43 terabytes of data in 72 hours cost less than \$2,000. With ADAPT in demand by other big customers, “what we processed over the weekend in AWS would have taken us months here in ADAPT given the available resources,” said Garrison Vaughan, NCCS systems engineer.

This summer the team will be re-processing the entire dataset while adding a layer of multispectral data. “The next one will have three times the accuracy with only about twice the overhead,” Tucker said. Stay tuned for news about this effort in Part 2.

Jarrett Cohen, NASA Goddard Space Flight Center

Contacts

Compton Tucker
Senior Physical Scientist
Earth Sciences Division
NASA Goddard Space Flight Center
compton.j.tucker@nasa.gov
301.614.6644

John David
Senior Programmer/Analyst
Earth Sciences Division
NASA Goddard Space Flight Center
john.l.david@nasa.gov
301.614.5737

More Information

Advanced Data Analytics Platform (ADAPT)
<http://www.nccs.nasa.gov/services/adapt>

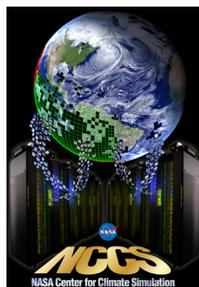


The NASA Center for Climate Simulation (NCCS) Advanced Data Analytics Platform (ADAPT) processed 260,000 satellite images to create mosaics for analysis on the commercial Amazon Web Services cloud. Photo by Jarrett Cohen, GSFC.

Hoot Thompson
Advanced Technology Lead
NASA Center for Climate Simulation
NASA Goddard Space Flight Center
john.h.thompson@nasa.gov
301.286.8567

Garrison Vaughan
Systems Engineer
NASA Center for Climate Simulation
NASA Goddard Space Flight Center
garrison.r.vaughan@nasa.gov
301.286.6283

This is How You Count All the Trees on Earth
http://www.theregister.co.uk/2015/12/09/nasa_amazon_supercomputing_trees_biomass_climate



NASA Center for Climate Simulation

High-Performance Science
<http://www.nccs.nasa.gov>