Earth Observation in the Cloud

AMS / OFCM Forum
Observing the Environment from the Ground Up

Jed Sundwall, AWS Open Data Global Lead

9 March 2016

Why does AWS care about open data?

esri THE CLIMATE CORPORATION ZOONIVERSE

Many of our customers rely on quality open data as much as they rely on our computing, storage, and other web services.

Traditional data acquisition

"…data must be organized, well-documented, consistently formatted, and error free. Cleaning the data is often the most taxing part of data science, and is frequently 80% of the work."
— *Data Driven* by DJ Patil and Hilary Mason

We ask: **How can we get rid of that 80%?**

Bandwidth and infrastructure constraints slow down research and development. New and better sensors are exacerbating this problem.

Data acquisition in the cloud

The cloud allows researchers to take their algorithms to data rather than downloading data to their computing resources.

When data is shared in the cloud, anyone can analyze it without needing to download it or store it themselves.

Data acquisition in the cloud

"Ordinarily, hitting 'copy' on a 4 gigabyte file is an opportunity to stand up and get a fresh cup of coffee, browse the sports section for a little while, but moving data between servers in an Amazon data center barely affords time to touch your toes a couple times."
— Paul Ramsey, describing Landsat on AWS

Public Data Sets on AWS

Several high-value datasets are available for anyone to access for free on AWS. Examples include:


3K Rice Genome


Landsat on AWS

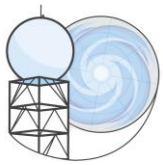

NEXRAD on AWS

Ask me about Sentinel-2 data and global elevation data if interested.

NEXRAD on AWS

The Next Generation Weather Radar (NEXRAD) is a network of 160 high-resolution Doppler radar sites that detects precipitation and atmospheric movement and disseminates data in 5 minute intervals from each site.

It has traditionally been time consuming and expensive to acquire, store, and analyze NEXRAD data. Accessing the full historical archive on demand has been almost impossible.



Learn more at: http://aws.amazon.com/noaa-big-data/nexrad/

7

NEXRAD on AWS

NEXRAD on AWS makes 270TB of individual volume scan files and real-time chunks as objects on Amazon S3.

Data can be accessed programmatically via a RESTful interface and quickly deployed to any of our products for analysis and processing.

Amazon Simple Notification Service allows subscription to notifications of new data.

8

NEXRAD on AWS

Before 2015

Step 1: Acquire 270TB of NEXRAD data.

Step 2: 😩😩😩

Step 3: Think of something else to do.

After 2015

Step 1: s3://noaa-nexrad-level2

Step 2: 🚀 The sky's the limit.

> "At some point, human intelligence became collective and cumulative in a way that happened to no other animal."
>
> — Prologue to *The Rational Optimist* by Matt Ridley

9

NEXRAD on AWS: Early Use Cases

- Climate Corporation cut two weeks off their development time for analysis and are incorporating the real-time feed into their production workflows.
- A weather data company stopped storing their own NEXRAD archive, freeing up revenue to build new products.
- Several weather companies developing new products based on the archival data and real-time feed.
- Unidata has made the archive data available via an AWS-hosted THREDDS Data Server to users with .edu domains.
- Several researchers are planning longitudinal analyses using the full archive.
- An open sensor data standards group is evaluating the NEXRAD on AWS model.

10

Thank you!

Jed Sundwall, AWS Open Data Global Lead
jed@amazon.com