



Location Powers: Data Science

Repeatable Science on Top of Oceans of
Shared Data

Mark Korver – Geospatial Lead, Specialist Team
Amazon Web Services

November 13-14, 2009



LOCATION POWERS

Open Source Foundational Tools

PostGIS



*Spatial data type
for PostgreSQL*

P R Ø J



GDAL

AWS News Blog

Amazon S3 – The First Trillion Objects

by [Jeff Barr](#) | on 12 JUN 2012 | [Permalink](#) | [Share](#)

Late last week the number of objects stored in Amazon S3 reached one trillion (1,000,000,000,000 or 10^{12}). That's 142 objects for every person on Planet Earth or 3.3 objects for every star in our Galaxy. If you could count one object per second it would take you 31,710 years to count them all.

We knew this day was coming! Lately, we've seen the object count grow by up to 3.5 billion objects in a single day (that's over 40,000 new objects per second).

Our customers have taken advantage of S3's relatively new [object expiration feature](#) and have used it to delete over 125 billion objects since we released it at the end of last year. In other words, even though we've made it easier to delete objects, the overall object count has continued to grow at a very rapid clip.

Resources

- [Getting Started](#)
- [What's New](#)
- [Top Posts](#)
- [Official AWS Podcast](#)
- [Case Studies](#)

Follow

AWS News Blog

Amazon S3 – Two Trillion Objects, 1.1 Million Requests / Second

by [Jeff Barr](#) | on 18 APR 2013 | [Permalink](#) | [Share](#)

Last June I blogged about the [first trillion objects](#) stored in Amazon S3. On the first day of re:Invent I updated that [number to 1.3 trillion](#).

It is time for another update!

I'm pleased to announce that there are now more than 2 trillion (2×10^{12}) objects stored in [Amazon S3](#) and that the service is regularly peaking at over 1.1 million requests per second.

It took us six years to grow to one trillion stored objects, and less than a year to double that number.

What Does That Mean?

It is always fun to try and put these numbers into real world terms:

Resources

- [Getting Started](#)
- [What's New](#)
- [Top Posts](#)
- [Official AWS Podcast](#)
- [Case Studies](#)

Follow

 [Twitter](#)

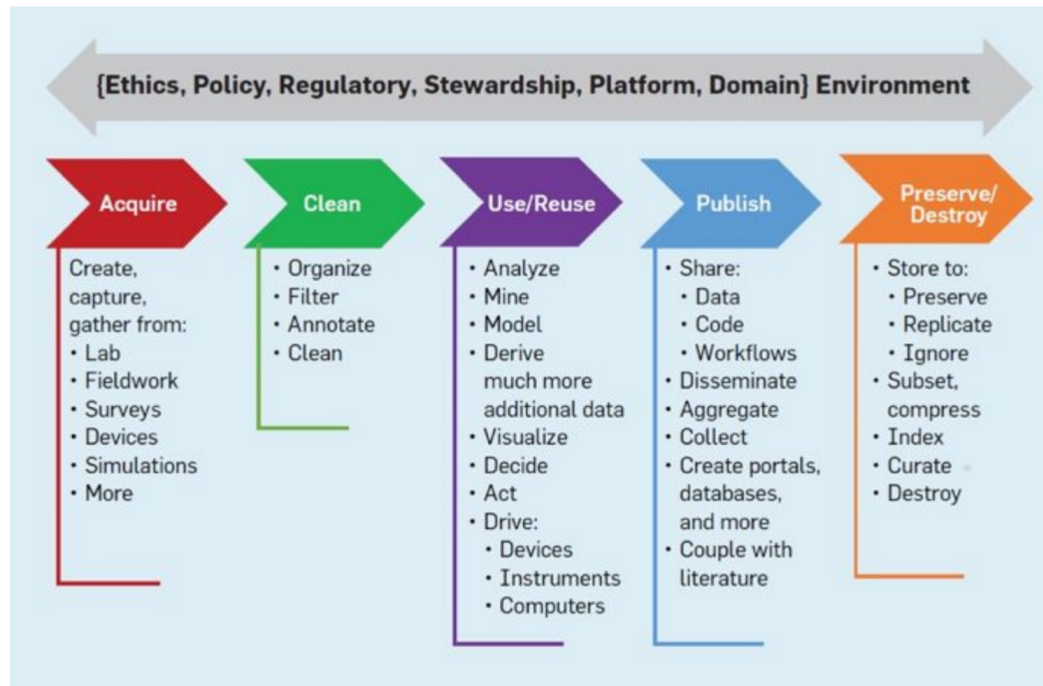


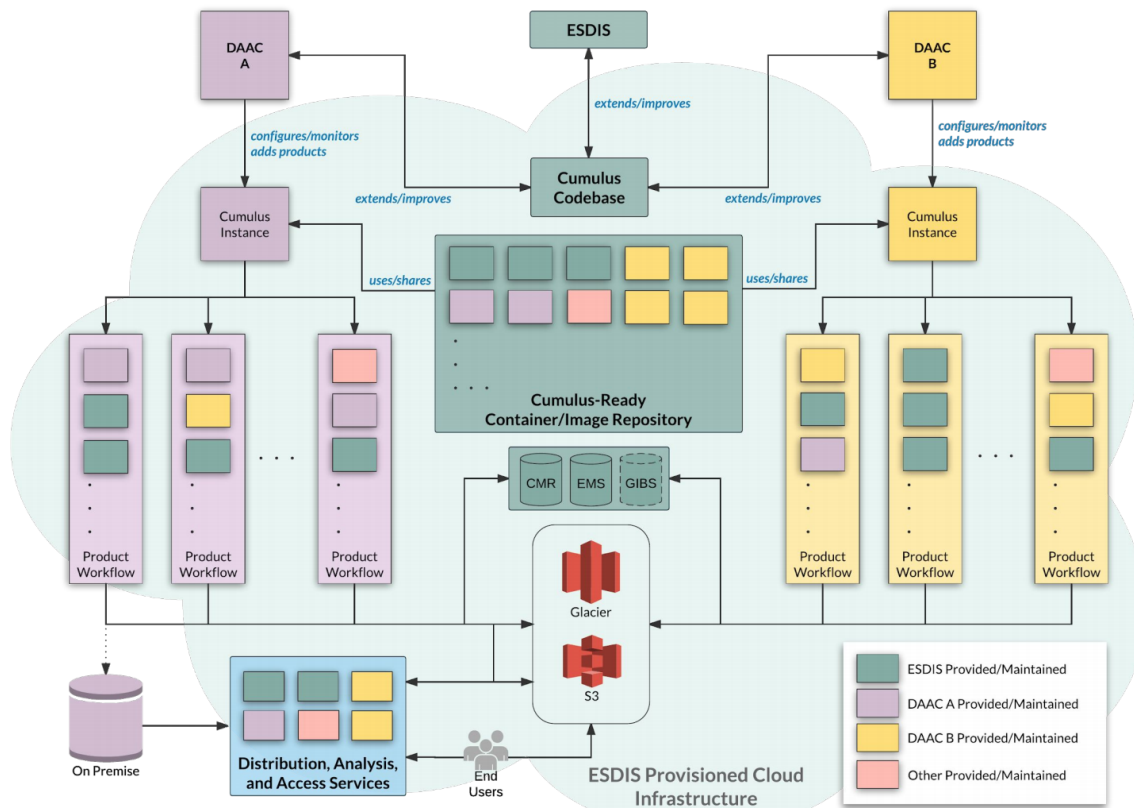
Figure. Data life cycle from the *Realizing the Potential of Data Science Report*.²

Methods in Data Science and defined in more limited discussions focuses on the mathematical and computer science algorithm-based techniques [Foundations of Data Science]

- o High-Dimensional Space

Application Silos

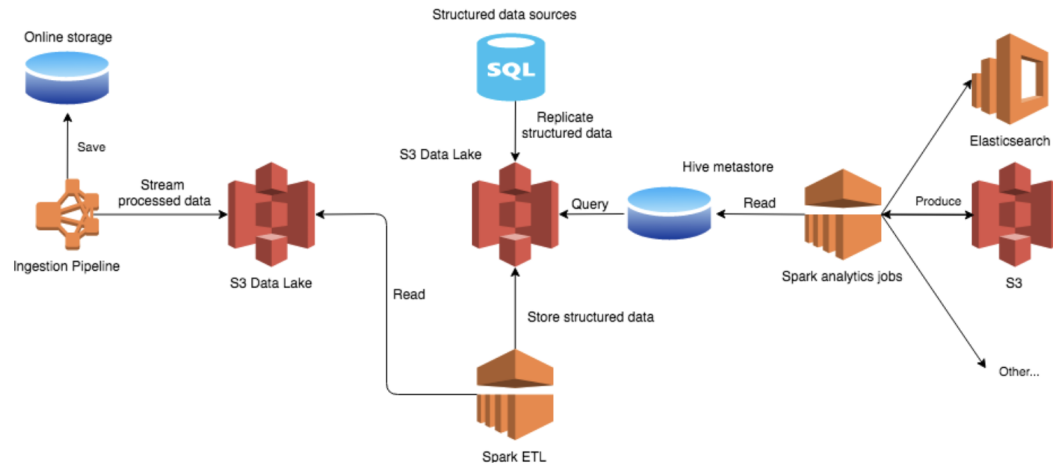




EOSDIS Cumulus Project - Earthdata - NASA

7

Here is how our data lake infrastructure roughly looks:



- The activity ingestion pipeline processes data and stores the data in online storage. Simultaneously, the pipeline streams data to a Kinesis

COG - Cloud Optimized GeoTIFF generator

Driver capabilities

Creation Options

File format details

Examples

See Also

COSAR - TerraSAR-X Complex SAR Data Product

CPG - Convair PolGASP data

CTable2 - CTable2 Datum Grid Shift

CTG - USGS LULC Composite Theme Grid

DAAS (Airbus DS Intelligence Data As A Service driver)

DB2 raster

DDS - DirectDraw Surface

DERIVED - Derived subdatasets driver

DIMAP - Spot DIMAP

DIPEX - ELAS DIPEX

DODS - OPeNDAP Grid Client

DOQ1 - First Generation USGS

[GDAL documentation](#) » [Raster drivers](#) » COG - Cloud Optimized GeoTIFF generator

[Edit on GitHub](#)

Previous

Next

COG - Cloud Optimized GeoTIFF generator

New in version 3.1.

Driver short name

COG

This driver supports the creation of Cloud Optimized GeoTIFF (COG)

It essentially relies upon the [GTiff - GeoTIFF File Format](#) driver with the `COPY_SRC_OVERVIEWS=YES` creation option, but automatically does the needed preprocessing stages (reprojection if asked and creation of overviews on imagery and/or mask) if not already done, and also takes care of morphing the input dataset into the expected form when using some compression types (for example a RGBA dataset will be transparently converted to a RGB+mask dataset when selecting JPEG compression)

Driver capabilities

Supports CreateCopy()

NEW FEATURES IN HDF5 RELEASE 1.12

HDF5

[Expand all](#) [Collapse all](#)

- › Learning HDF5
- › HDF5 Examples
- › HDF5 User's Guide
- ✓ HDF5 Application Developer's Guide
 - ▼ Release Specific Information
 - › **New Features in HDF5 Release 1.12**
 - Software Changes from Release to Release for HDF5-1.12
 - Migrating from HDF5 1.8 to HDF5 1.10
 - › New Features in HDF5 Release 1.10
 - › Software Changes from Release to Release for HDF5-1.10
 - New Features in HDF5 Release 1.8
 - Software Changes from

⚠ This release includes changes in the HDF5 storage format. **PLEASE NOTE that HDF5-1.10 and earlier releases cannot read files created with the new features described below that are marked with a ***.

HDF5 1.12 introduces several new features in the HDF5 library:

- [Virtual Object Layer \(VOL\)](#) *
- [Virtual File Drivers - S3 and HDFS](#) *
- [Hyperslab Performance Improvement](#)
- [Update to References](#) *
- [Update to Selections](#)

Virtual Object Layer (VOL) (RFC) *

See the [Virtual Object Layer](#) page for more information.

The Virtual Object Layer (VOL) is an abstraction layer within the HDF5 library that enables different methods for accessing data and objects that conform to the HDF5 data model. The VOL intercepts all HDF5 API calls that potentially modify data on disk and forwards those calls to a plugin "object driver". The data on disk can be a different format than the HDF5 format.


Announcing the SpatioTemporal Asset Catalog (STAC) specification




Chris Holmes [Follow](#)

Nov 29, 2017 · 7 min read

Today I am pleased to announce that the SpatioTemporal Asset Catalog (STAC) repository is 'open for business'. This was the result of a lot of work by many amazing people at the Boulder sprint, bringing together a huge variety of perspectives to increase interoperability in searching for satellite imagery and other spatiotemporal assets. There is still lots of work to do to turn STAC into a really solid specification, but we invite any interested developer to check out the latest version, implement in their software, and participate in the open collaboration to iteratively improve it.

 This repository Search Pull requests Issues Marketplace Explore

 radiantearth / stac-spec

Unwatch 18

Star 8

Fork 11

[Code](#) [Issues 9](#) [Pull requests 0](#) [Projects 0](#) [Insights](#) [Settings](#)

Registry of Open Data on AWS



About

This registry exists to help people discover and share datasets that are available via AWS resources. [Learn more about sharing data on AWS.](#)

See [all usage examples for datasets listed in this registry](#) tagged with sustainability.

Search datasets (currently 58 matching datasets)

You are currently viewing a subset of data tagged with sustainability.

Add to this registry

If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the [Registry of Open Data on AWS GitHub repository](#).

Unless specifically stated in the applicable dataset

Sentinel-2

disaster response earth observation geospatial natural resource
satellite imagery sustainability

The [Sentinel-2 mission](#) is a land monitoring constellation of two satellites that provide high resolution optical imagery and provide continuity for the current SPOT and Landsat missions. The mission provides a global coverage of the Earth's land surface every 5 days, making the data of great use in on-going studies. L1C data are available from June 2015 globally. L2A data are available from April 2017 over wider Europe region and globally since December 2018.

[Details](#) →

Usage examples

- [Exploring the Chile wildfires with Landsat and Sentinel-2 imagery](#) by Timothy Whitehead
- [Integrate imagery from the Sentinel-2 archive into your own apps, maps, and analysis with the Sentinel-2 image service](#) by Esri
- [Using Vector tiles and AWS Lambda, we can build a really simple API to get Landsat and Sentinel images](#) by Remote Pixel
- [EOS Land Viewer](#) by Earth Observing System
- [Sentinel Playground](#) by Sinergise

[See 17 usage examples](#) →