Earth Observation in the Cloud

Esri Canada User Conference – Ottawa

Jed Sundwall, Open Data

6 October 2015

amazon web services

Agenda

- What is cloud computing?
- Working with data on AWS
- Open data on AWS
- Lessons from Landsat on AWS
- Esri on AWS

What is cloud computing?

What is cloud computing?

Cloud computing is the on-demand delivery of IT resources via the Internet with pay-as-you-go pricing.

Why Are Customers Adopting Cloud Computing?

Seven main benefits experienced by customers in virtually every industry

1. It replaces up-front capital expense with low variable cost.
2. It offers lower variable costs than companies can achieve themselves.
3. It provides pricing-model choice to support variable and stable workloads.
4. It drives down IT labor costs, both up-front and ongoing.
5. It offers a premium security spec at non-premium prices.
6. It supports highly available workloads for a fraction of the cost.
7. It saves you more money as you grow bigger.

AWS has announced price reductions 49* times since our inception in 2006. Recent price drops included…

34%

Amazon ElastiCache reduces prices for cache nodes by an average of 34%

March 26, 2014

51%

Amazon S3 reduces prices for Standard and Reduced Redundancy Storage, by an average of 51%

March 26, 2014

20%

Amazon Route 53 lowers prices for both standard queries and latency-based routing queries by 20%

July 31, 2014

* As of June 2015

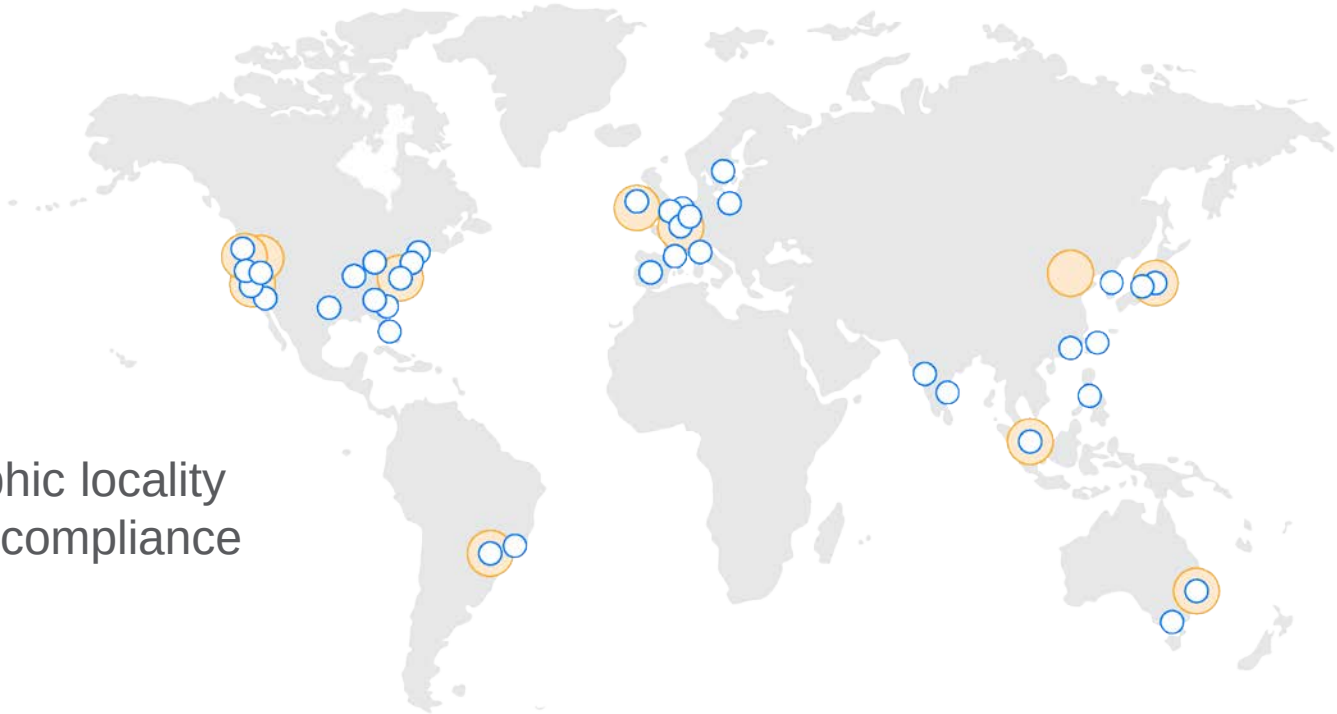
AWS Global Infrastructure

11 Regions

28 Availability Zones

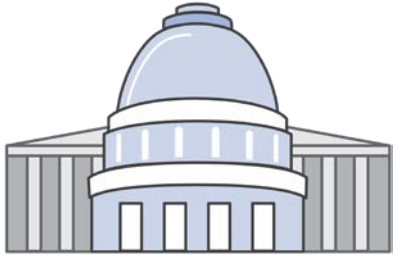
53 Edge locations

Control your geographic locality
for performance and compliance



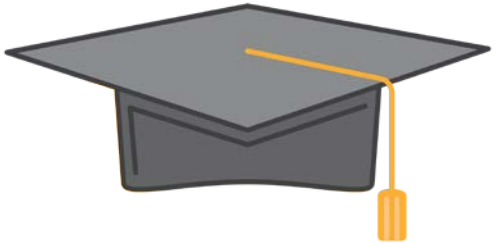
amazon web services

AWS in the Public Sector



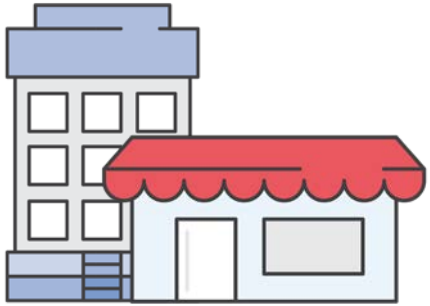
1,700+

government agencies



4,500+

educational institutions



17,000+

nonprofit organizations

Government Agencies and Educational Institutions Use AWS Worldwide



















amazon web services

Working with data on AWS

Data on AWS

Amazon Web Services provides a comprehensive toolkit for gathering, storing, analyzing, and working with data at any scale.

Amazon S3 lets you store and retrieve any amount of data, at any time, from anywhere on the web.

Amazon Elastic MapReduce (Amazon EMR) provides the Apache Hadoop analytics framework as an easy-to-use managed service.

Amazon DynamoDB is a fully-managed NoSQL database service that makes it cost-effective to store and retrieve any amount of data.

Amazon S3

Highly durable object storage for all types of data

Internet-scale storage
Grow without limits

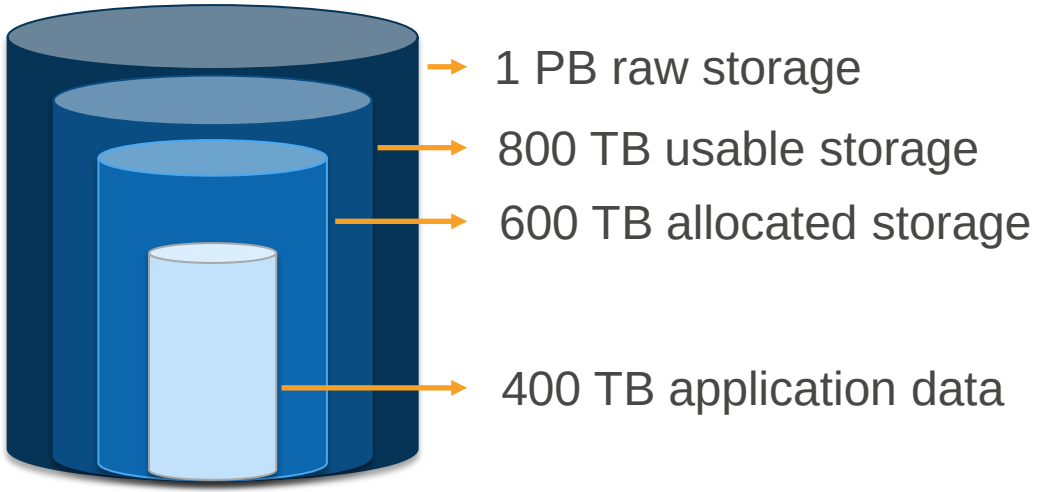
Built-in redundancy
Designed for 99.999999999% durability

Low price per GB per month
No commitment
No up-front cost

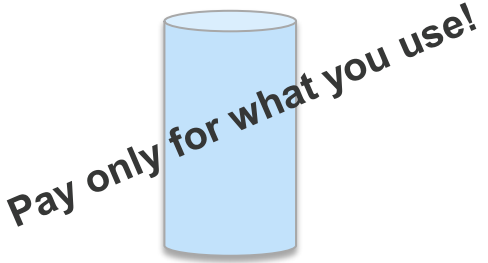
Benefit from AWS's massive security investments

Amazon S3

Pricing at http://aws.amazon.com/s3/pricing/

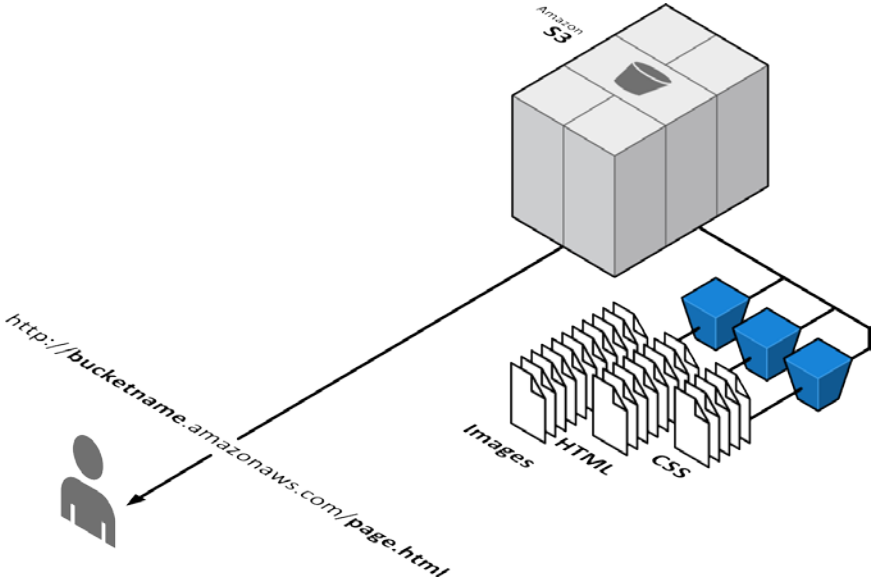


Traditional Storage



Amazon S3

Serving files with Amazon S3



Amazon S3 can serve files over the Internet.

Collaborators anywhere in the world can access data stored in Amazon S3 to perform analysis, or create databases.

Open data on AWS

What is open data?

Open data is data that can be used by anyone for any purpose for free.

Many of our customers, such as Esri, the Weather Company, and the Climate Corporation, rely on quality open data as much as they rely on our computing, storage, and other web services.

The Weather Company saves $1 million per year running its forecasting application on AWS

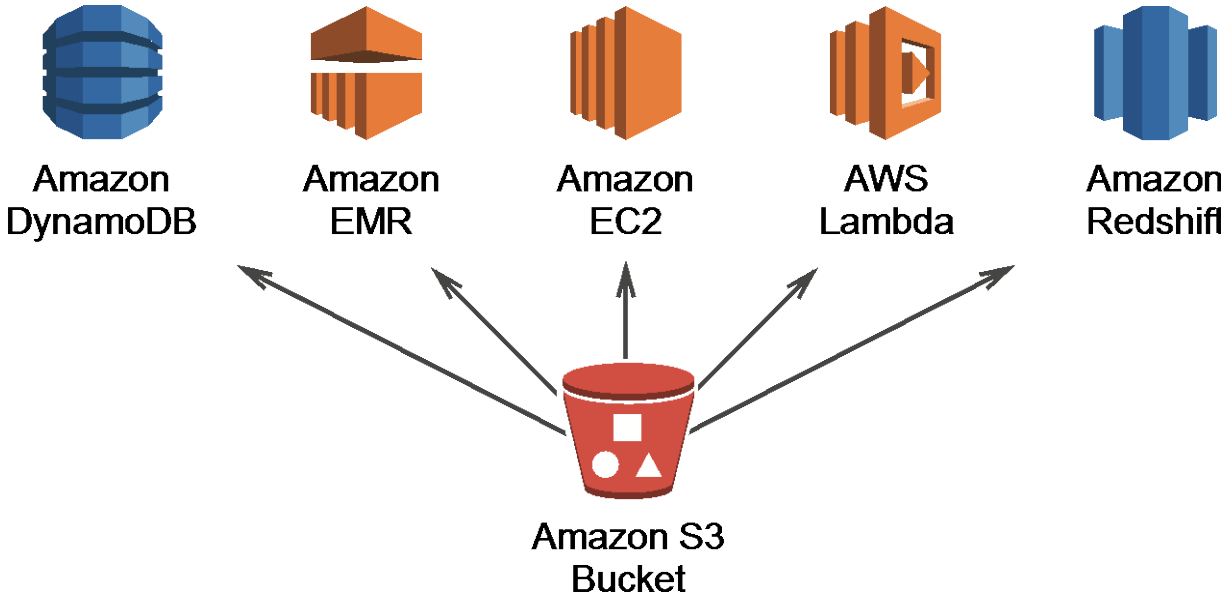
> Using AWS, TWC can scale as necessary to handle constantly changing workloads and maintain our 11-millisecond response time.
>
> **Bryson Koehler**
> EVP, CTO, CIO, The Weather Company

The Weather Company provides millions of people with the world's best weather forecasts, content and data, every day.

- Needed a cost-effective, scalable alternative to operating 13 data centers with legacy systems.
- TWC ingests, stores, and analyzes 4 GB of weather data per second from over 800 sources.
- Designed to handle more than 15 billion API calls each day, at a rate of 150,000 per second.
- Reduced its on-premises IT environment form 13 to 6 data centers.

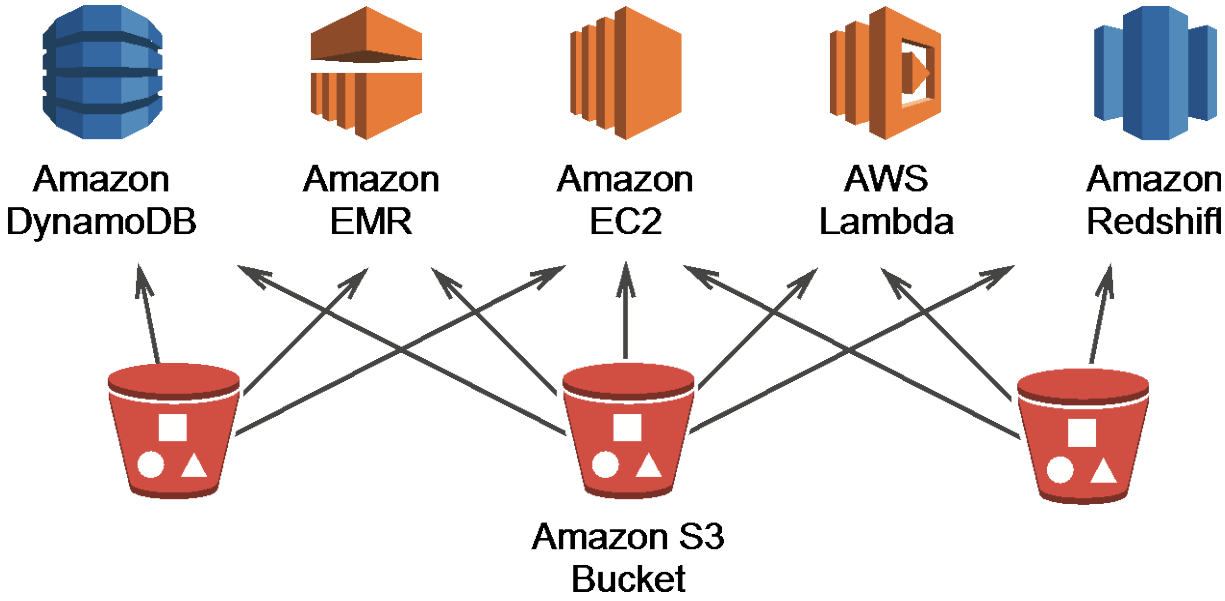
The power of open data on AWS

Making data open on AWS enables more innovation by making data available for rapid access to our flexible and low-cost computing resources.

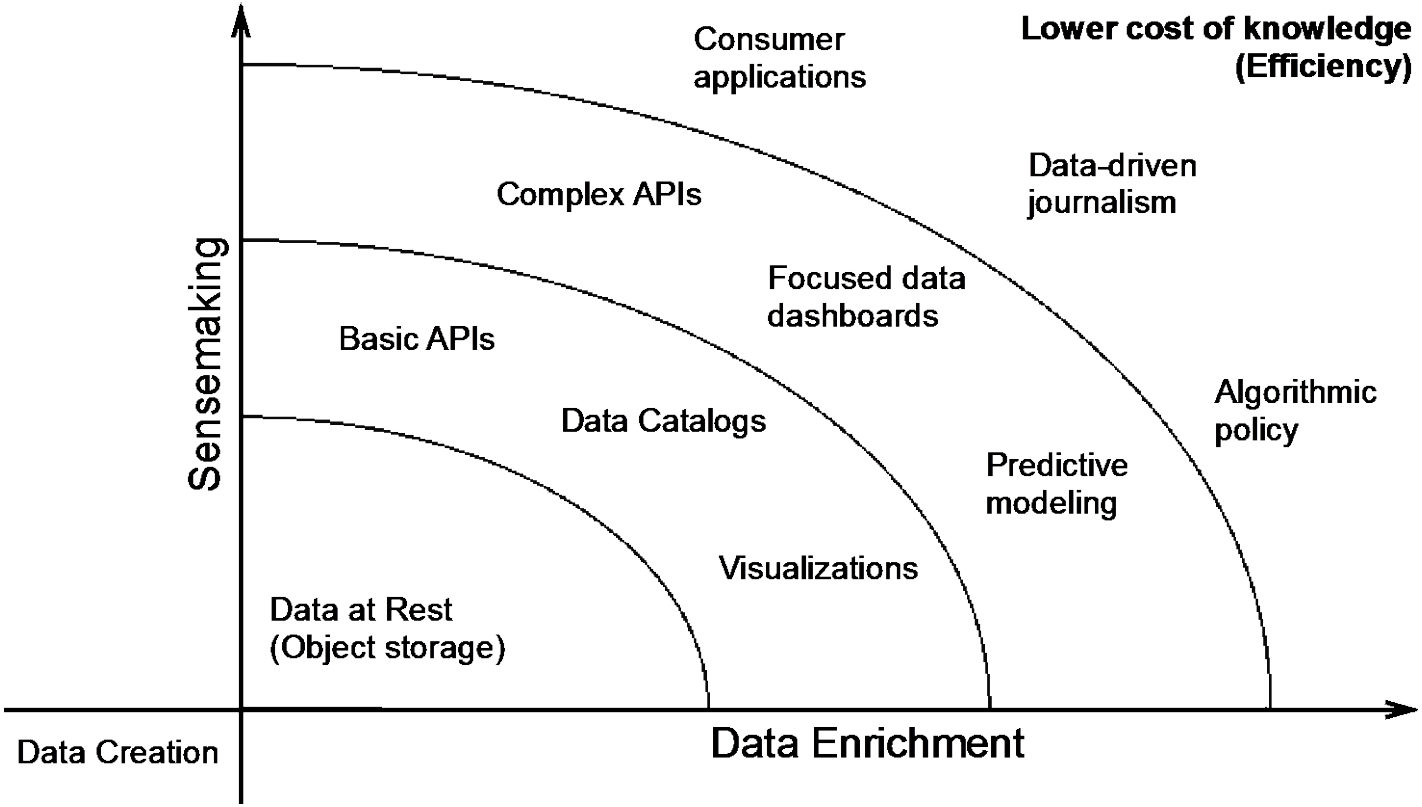


The power of open data on AWS

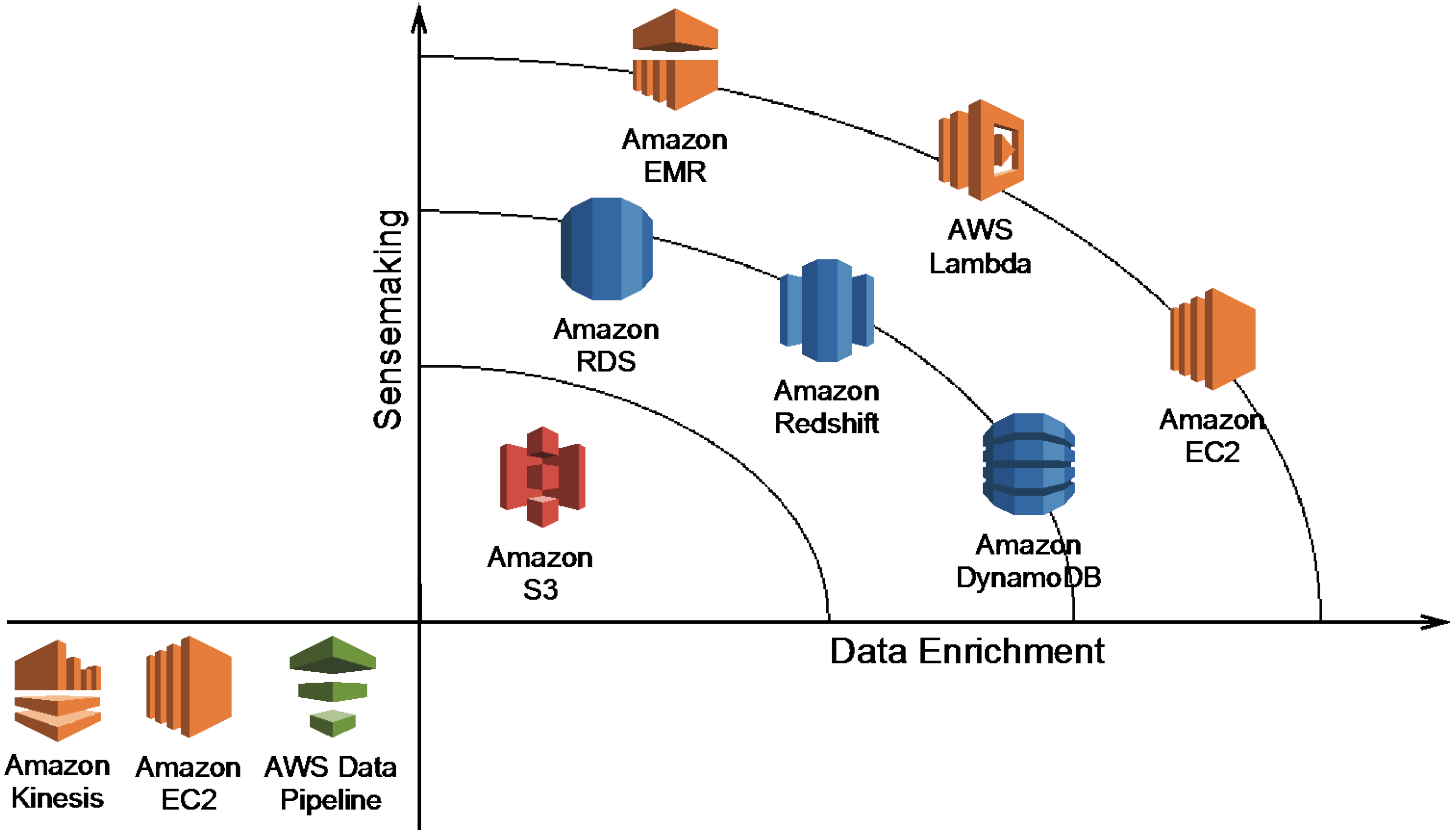
Making data open on AWS enables more innovation by making data available for rapid access to our flexible and low-cost computing resources.



Open data as a platform



Open data as a platform



An Amazonian approach to open data

Two ideas that inform how we approach public data sets:

- Work backwards from the customer
- Eliminate undifferentiated heavy lifting

Working Backwards

“To ensure that a service meets the needs of the customer (and not more than that) we use a process called “Working Backwards” in which you start with your customer and work your way backwards until you get to the minimum set of technology requirements to satisfy what you try to achieve. The goal is to drive simplicity through a continuous, explicit customer focus.”

—Werner Vogels, CTO of Amazon

Working Backwards

- Seek out valuable data by listening to customer needs
- Consider real-world use cases for the data
- Consider the size of the user community or market opportunity

Undifferentiated heavy lifting

"…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

Driving innovation with open data

To drive innovation with your data, you need to ensure three core things when sharing it:

1. The data is accurate
2. The data is provided with documentation
3. The data will still be available to developers tomorrow

Note: these aren’t necessarily technical problems.

Landsat on AWS

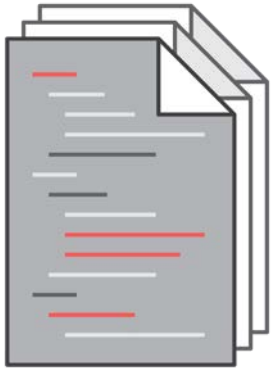
amazon web services

Public datasets on AWS

To enable more innovation, AWS hosts a selection of datasets that anyone can access for free. Data in our public datasets is available for rapid access to our flexible and low-cost computing resources.

Life Sciences
1000 Genomes Project

Earth Science
Landsat on AWS

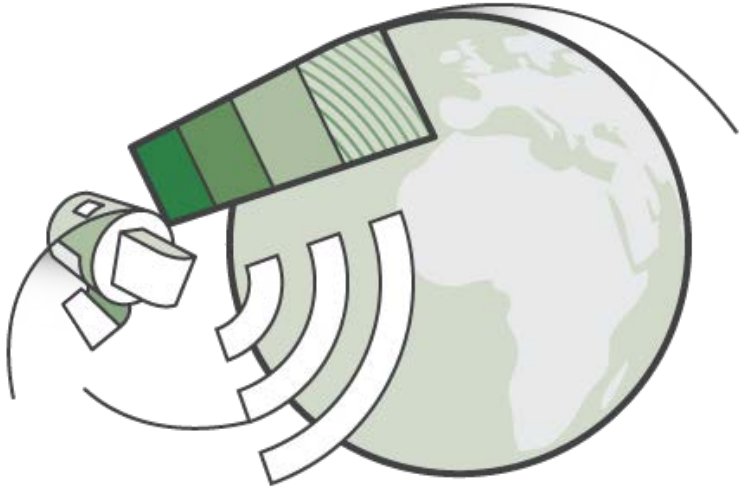


Internet Science
Common Crawl Corpus

Landsat on AWS

We have committed to making up to 1 petabyte of Landsat imagery readily available as objects on Amazon S3.

All Landsat 8 scenes from 2015 are available, along with a selection of cloud-free scenes from 2013 and 2014. All new Landsat 8 scenes are made available each day (~700 per day), often within hours of production.

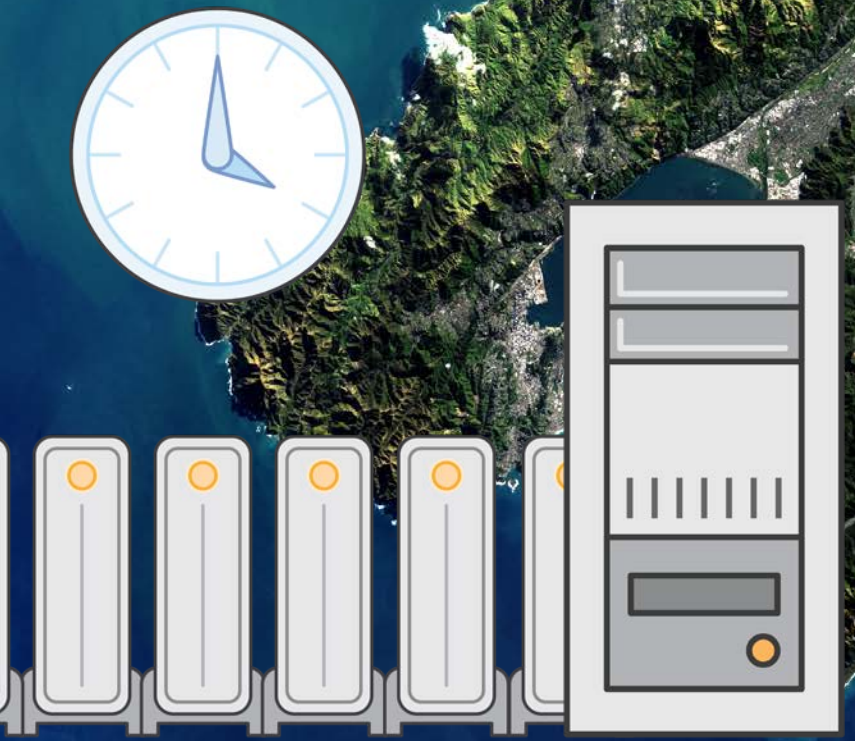


What is Landsat?

The Landsat program is a joint effort of the U.S. Geological Survey and NASA. It is the longest running program to gather Earth imagery from space and is considered the gold standard for natural resources satellite imagery.

amazon web services

Landsat is big open data



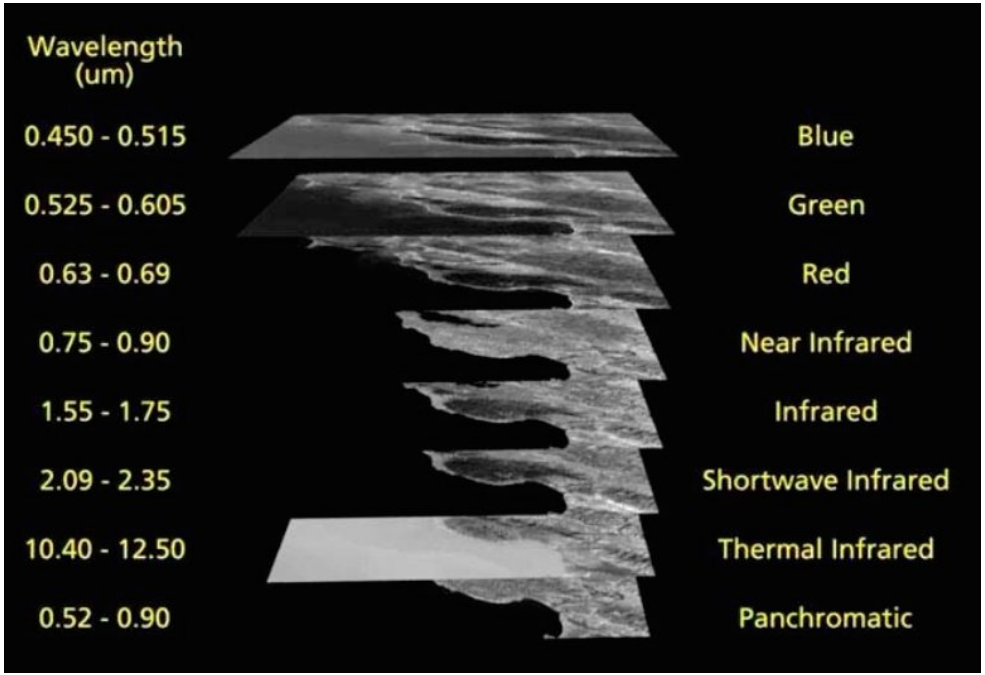
The Landsat program is a joint effort of the U.S. Geological Survey and NASA. It is the longest running program to gather Earth imagery from space and is considered the gold standard for natural resources satellite imagery.

It has traditionally been time-consuming and expensive to acquire, store, and analyze Landsat data.

What is Landsat?

Landsat scenes are made up of multiple files, each of which includes data about different kinds of light reflected off of Earth.

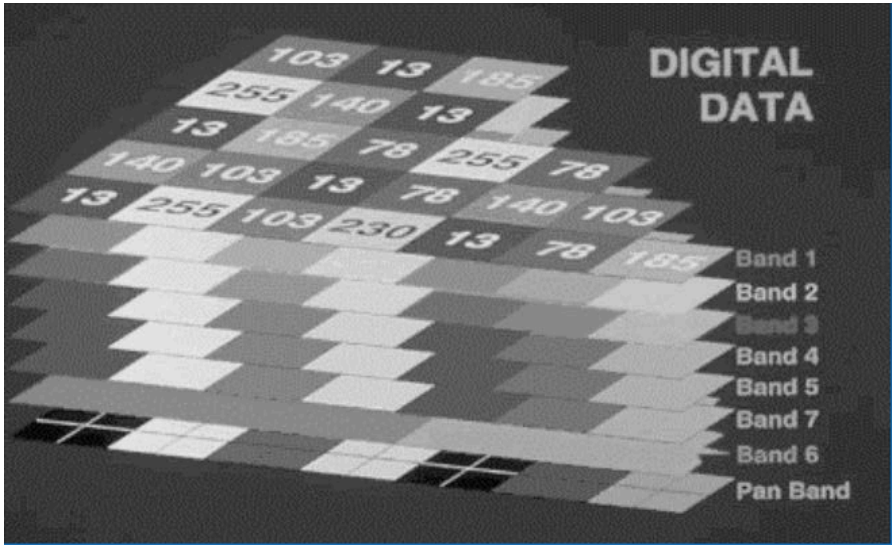
The Landsat program has been in operation since 1972.



What is Landsat?

Each pixel of each Landsat 8 file represents a 12-bit measurement of light reflected off a 30m^2 part of our planet.

Each Landsat 8 scene contains about 840 million pixels and takes up about 800 MB.

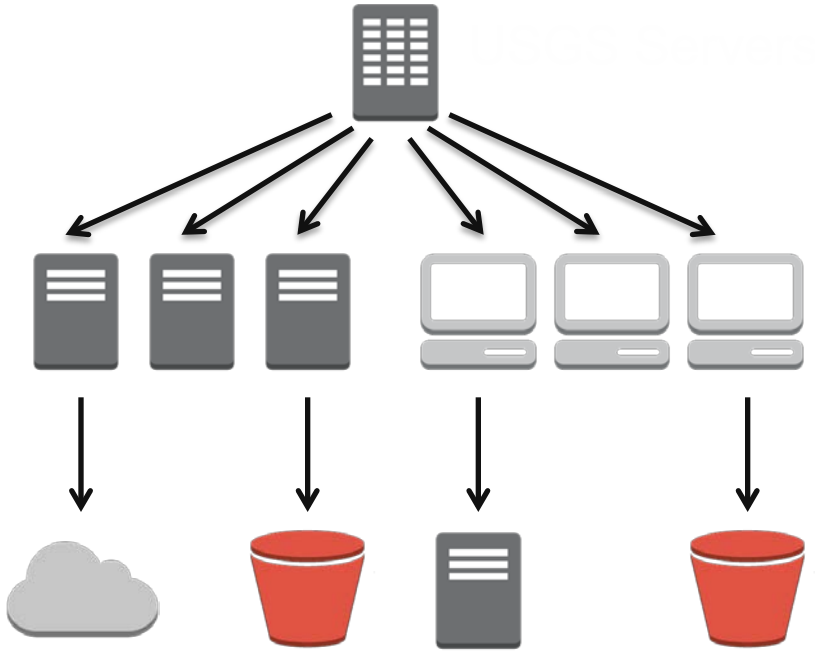


The Traditional Approach

Data is most commonly accessed via a web interface and downloaded on premises before being loaded into a web server.

All bands are downloaded in a .tar archive, even if you only need a few bands.

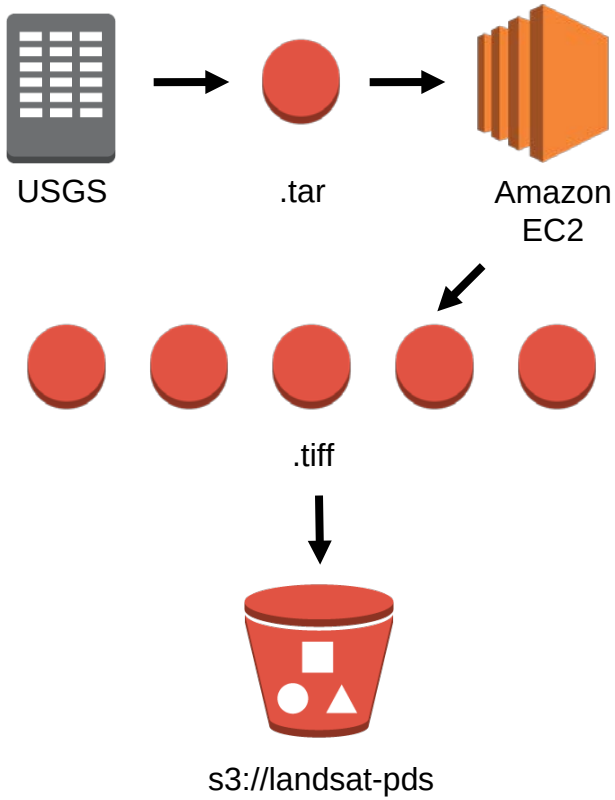
Data acquisition is time consuming and inherently redundant. Analysis is limited by user's access to bandwidth, storage, memory, and processing power.



Landsat on AWS

Landsat on AWS makes each band of each scene readily available as objects on Amazon S3. They can be accessed programmatically via HTTP and quickly deployed to any of our products for analysis and processing.

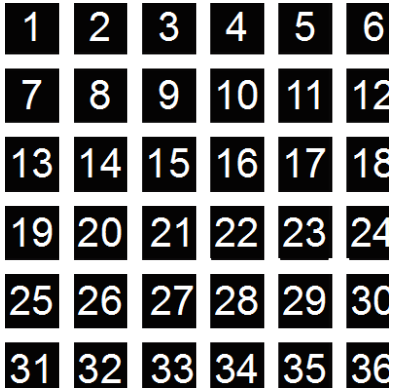
Users do not need to worry about local storage and have access to virtually unlimited computing power on demand.



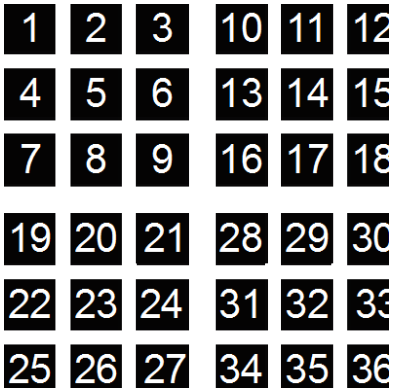
Undifferentiated heavy lifting

We use GDAL to add “internal tiling” on each Landsat on AWS tiff, which allows developers to use HTTP range gets to access specific portions of each scene.

This allows people to only access the data they need when they need it.



Standard tiff object



Internal tiled tiff object

Think of URLs instead of copies

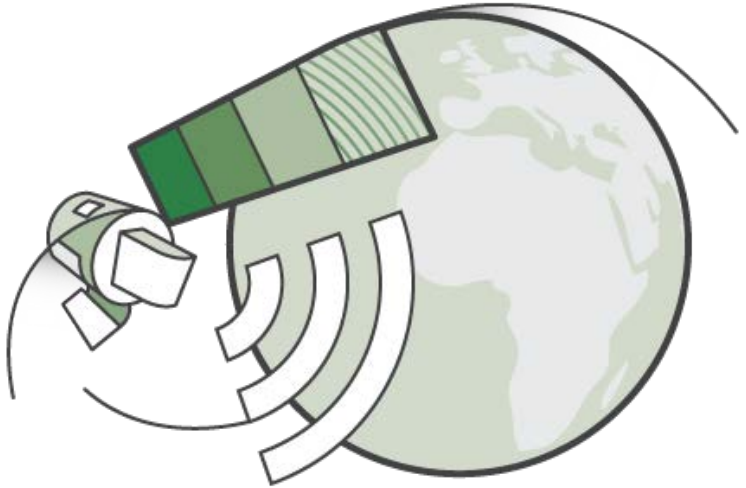
Wellington, New Zealand
https://landsat-pds.s3.amazonaws.com/L8/072/089/


Wellington, New Zealand – June 29 2015
RGB
Visible light
Infrared
Vegetation
Shortwave infrared
Urban areas
amazon
web services

Landsat on AWS

In the first 150 days (March 19 to August 16)

- Over 200,000 scenes available
- 3.8 PB of data accessed
- Over 560 million hits from 167 countries
- 3.7 million average requests per day
- 25 TB average data transferred per day



Snapsat

A team of five novice programmers used Landsat on AWS to build a web app called Snapsat that creates Landsat data visualizations in seconds.

Snapsat was built during the team's 8-week training program at Code Fellows. They launched it just months after learning to write code.

http://snapsat.org

Esri – Unlock Earth's Secrets

Esri has created a tool to show how ArcGIS Online can quickly visualize Landsat data for live visualization and analysis within the browser.

"These are not pre-generated cache services limited to just visualization—they are dynamic, high-performance image services that perform on-the-fly processing and dynamic mosaicking of Landsat's multi-spectral and multi-temporal imagery."

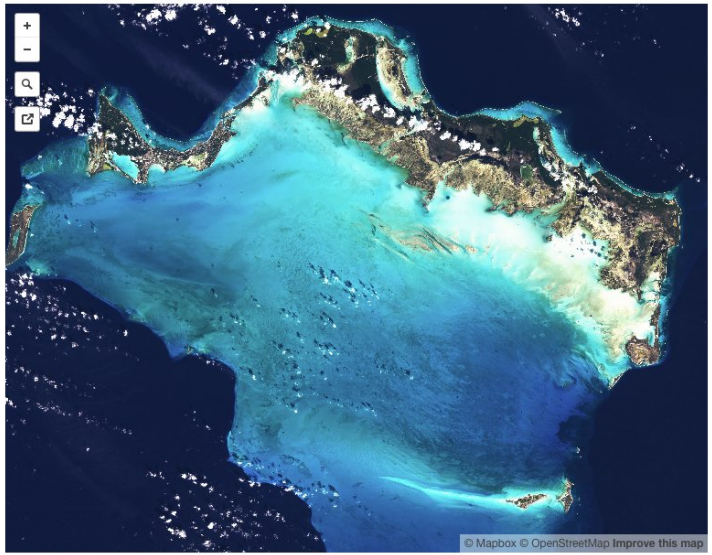
http://www.esri.com/landsatonaws

amazon web services

landsat-util

Landsat on AWS helped Development Seed make optimizations that make landsat-util over 2× faster and allow for more functionality.

```
landsat search --download --imageprocess --pansharpen  --cloud 4 --start "janua
ry 1 2014" --end "january 10 2014" pr 009 045
```



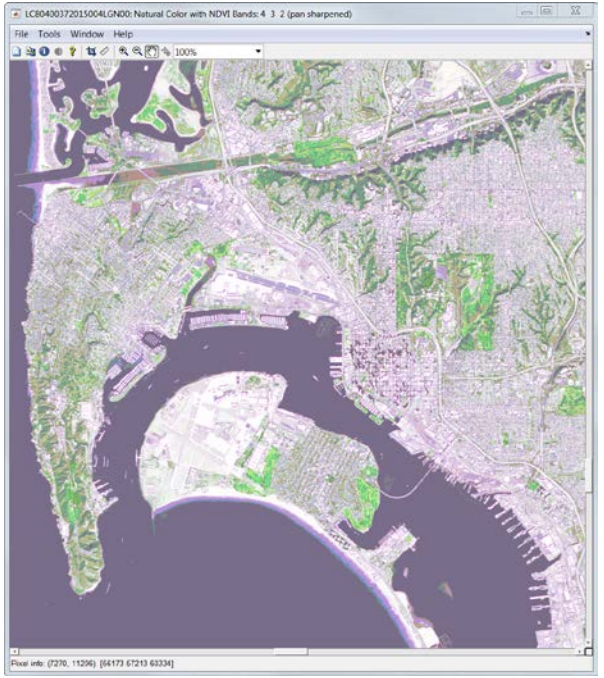
Turks & Caicos Islands, British West Indies

https://developmentseed.org/blog/2015/03/19/aws-landsat-archive/

MATLAB – Landsat8 Data Explorer

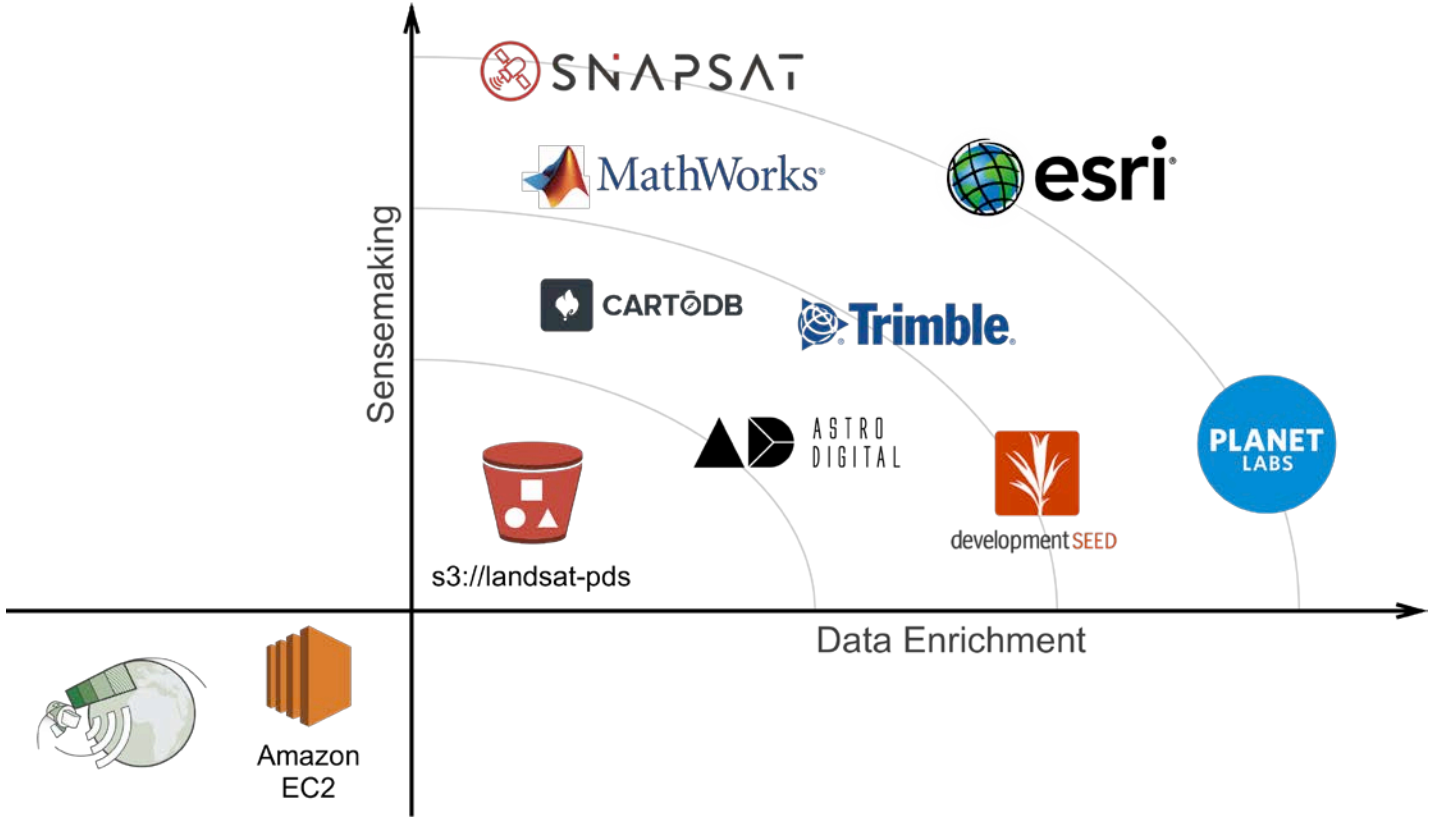
MathWorks created a freely downloadable MATLAB based tool for accessing, processing, and visualizing Landsat 8 data.

The tool allows MATLAB users to find Landsat 8 scenes, analyze them, and combine them with other sources of GIS data for new visualizations.

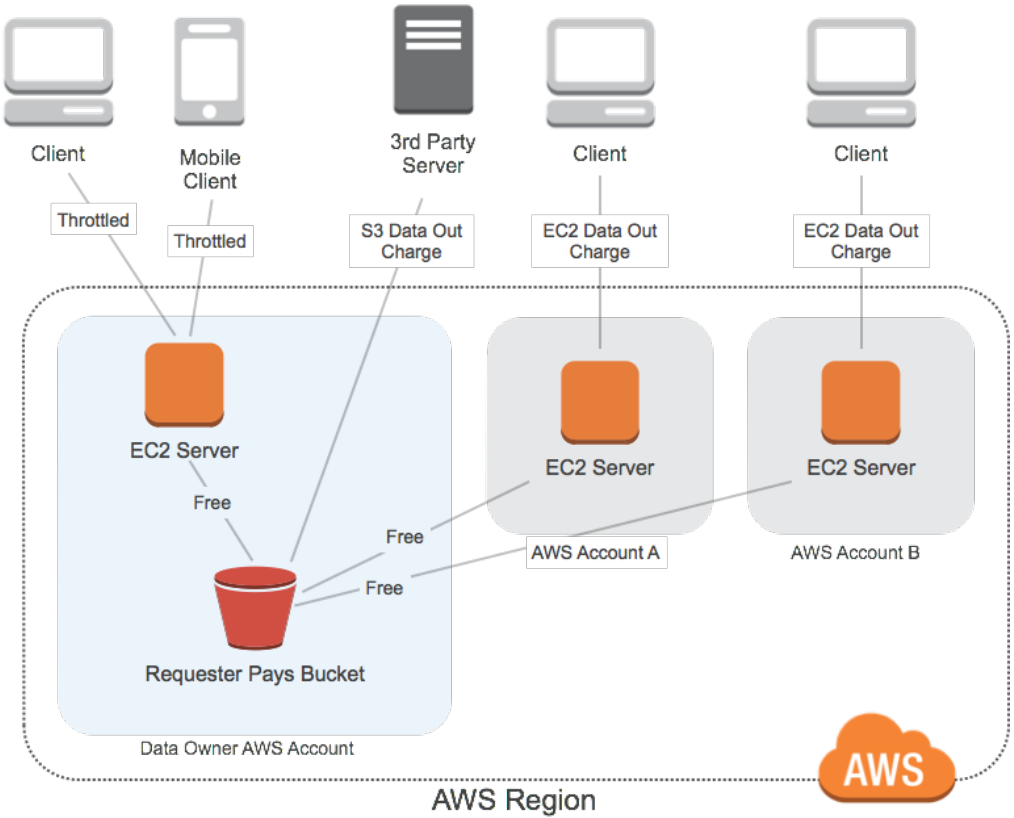


http://blogs.mathworks.com/steve/2015/03/19/matlab-landsat-8-aws/

Landsat on AWS as a platform



Shared data architecture



Esri on AWS

amazon web services

Esri on AWS

In a matter of minutes, you can have your ArcGIS for Server running on Amazon Elastic Compute Cloud (EC2).

Benefits include:

- Accelerated deployment of ArcGIS for Server
- Reduced hardware and software management requirements
- Self-service access to Amazon EC2 features, such as Auto-scaling and Elastic load balancing

Amazon EC2

http://www.esri.ca/en/content/arcgis-server-cloud

Esri OptimizeRasters

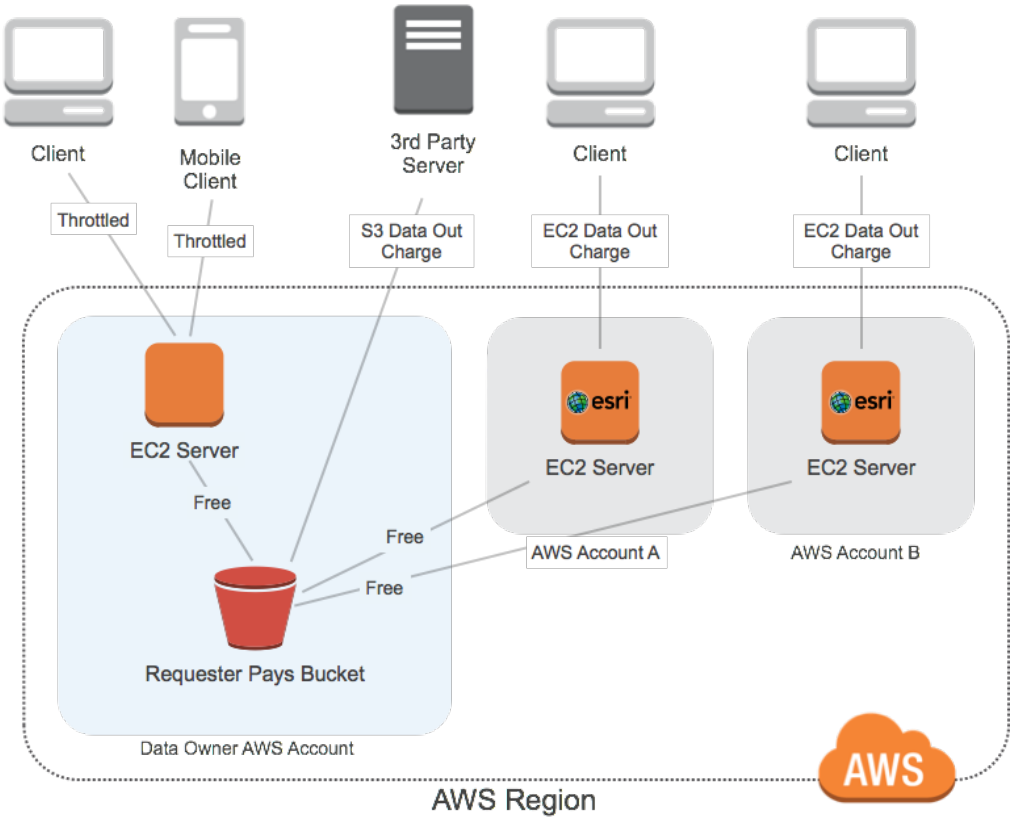
Esri's Peter Becker is developing new ways for ArcGIS Server to access and analyze imagery stored in the cloud.

This allows ArcGIS users to benefit from low-cost scalable storage of raster data.

Amazon S3

For more information, email ImageManagementWorkflows@esri.com

Shared data architecture with Esri



Thank you!

Jed Sundwall, Open Data

jed@amazon.com