Access to Cloud Raster Data Using GDAL, MRF and LERC

Peng Gao, Peter Becker, Thomas Maurer, Abhijit Doshi, Lucian Plesea
Esri, 380 New York Street, Redlands, CA 92373
lplesea@esri.com
Team Credentials

• Peng Gao
  - Esri, Development Lead, Imagery

• Peter Becker
  - Esri, Product Manager - Imagery

• Thomas Maurer
  - Esri Computer Scientist, image processing

• Abhijit Doshi
  - Esri, Product Engineer

• Lucian Plesea
  - Esri Computer Scientist, geospatial data services
The Esri approach to cloud enabling geospatial imagery and data services

- The Core Technology: Mosaic Dataset
- Landsat 8 hosted service
  - Initial implementation
  - Improvement by use of Meta Raster Format (MRF)
  - Improvement by use of LERC
- Landsat 8 cloud based service
  - Use of Object Stores (S3)
- Landsat 8 cloud service mashup
  - Built on Amazon storage grant
- Other cloud enabled Raster Data
High Level GIS Service: Mosaic Dataset and Image Services

- A mosaic dataset is a data model for management of large collections of imagery/rasters
- A mosaic dataset stores in a database:
  - References to image/rasters
  - Metadata about rasters
  - Processing to be applied
- Data selection, fusion of input rasters and any other processing is done on demand
- An image services is a dynamic web service which exposes the capabilities and content of a mosaic dataset
- Scalable to extremely large dataset collections
- Performance is dependent on the source data features and access

Access to Cloud Raster Data
Esri Landsat 8 Image Service

- Landsat 8 data is available free of charge
- Esri providing ArcGIS Online community access to recent Landsat 8 data
  - Global coverage, full resolution and spectrum
  - All scenes since 1/1/2015 and 5 best scenes of 2013,2014
  - Selected based on cloud coverage
  - Continuously updated as new scenes become available (approx. 400GB/day)
  - Common Landsat 8 processing functions are preconfigured
Optimizing the Response Time

- The initial implementation performance was lacking
  - About eight second average response time
  - Input format had significant latencies and read inefficiencies
  - Access to data on shared storage contributes to the delay
  - Web hosting with significant storage is not common and pricy

- Possible areas of improvement
  - Transcode the data into a format with faster access
  - Reduce the number of IO operations to slow storage
Limited Error Raster Compression (LERC)

- Esri algorithm for data compression
- Years of use in ArcGIS Server as a web transmission format
- Does not rely on sequence matching (like LZW, DEFLATE) nor on a space transform (Wavelet, DCT)
- Very fast decode and encode, about one order of magnitude faster than PNG
- Loss-less or lossy, based on user provided maximum error
  - Uses quantization internally
- Explicit data mask, making it efficient for sparse and projected swath raster data
- Patented but being released for geospatial application
Meta Raster Format (MRF)

- Raster format originated at JPL, for large raster datasets and web server tile
  - Open source GDAL driver
  - Abstracts tiling and pyramid organization
  - Various tile (subraster) compression formats
  - Separate and simple metadata, index and raster data components (files)
- The ability to located the different components on different classes of storage makes it a valuable technology for cloud GIS
Amazon Storage Options

- **Ephemeral** (80GB) $0 (inc. with EC2)
- **EBS (Elastic Block Storage)** 1TB/Disk ($100/TB/Month)
- **S3 Object Store** $30/TB/Month
  - 99.999999999% Durability
  - 99.99% Reliability/year
  - ≈10-50MB/s

- **NFS**
- **HTTP**
Landsat 8 Service, Second Hosted Implementation

- Customized ArcGIS Server that includes MRF+LERC
- Splitting MRF between ephemeral and EBS doubles the access speed vs. TIFF
- Using LERC compression doubles the access speed again vs. DEFLATE
- Average request latency reduced from eight to two seconds
  - Performance is very close to directly connected storage
  - Acceptable for interactive applications
- Added pre-processing
  - Trans-coding of data from downloaded TIF to MRF
  - Had to be implemented on SSD to achieve reasonable performance
- This implementation became public

Access to Cloud Raster Data
Landsat on AWS

- Landsat 8 data is available for anyone via Amazon S3
  - All 2015 scenes, selections of cloud free 2013-2014
  - Stored as geoTIF with Deflate compression and 3X pyramids

- Esri Cloud based Landsat 8 server
  - Based on the Amazon Landsat collection
  - S3 latency is large, comparable with a slow shared network file system
  - Data is converted to local MRF on demand, then cached locally and reused
Esri Landsat 8 Service: Mash-up

- [http://www.esri.com/landsatonaws](http://www.esri.com/landsatonaws)
- Fully cloud based, on AWS and S3
- Enabling technology is MRF with LERC compression, from S3 object store
- Response times are in the 3-4 seconds initially, dropping to 1-2 seconds for repeated access
- Easy to scale out since no local data is required initially
Other Esri cloud raster services: NAIP

- **USDA National Agriculture Imagery Program**
  - A primary goal of the NAIP program is to make digital ortho-photography available to governmental agencies and the public within a year of acquisition
  - Each state mosaic is redone every three years

- **Esri is collaborating with USDA to help achieve the timely release goal**
  - Service available on ArcGIS OnLine
  - Updated at the same time as new NAIP data is released
  - DOQs are converted to MRF and LERC
  - A Mosaic Dataset allows access to the whole NAIP collection

Data intensive cloud based GIS

- Reduces cost through collaboration
  - Open Source (GDAL, NASA/ESRI MRF)
  - Commercial contributions (Amazon, Esri)

- ArcGIS provides Server SW that runs in multiple cloud environments

- Use of MRF and LERC enables faster image processing and analysis of data stored in Object Storage
Understanding our world.