Distributed Image Analysis Using the ArcGIS API for Python

Jie Zhang, Elizabeth Graham
The ArcGIS Platform
Is a Comprehensive Imagery Platform

System of Insight
Extract Information from Imagery

Professional Imagery & Geospatial Analysts

System of Engagement
Manage and process all your imagery

System of Record

Content from:
ArcGIS Online, Partners, Your Org.
Enterprise Image Server with Distributed Raster Analysis

- **From 10.5, ArcGIS has a new way to create and execute spatial analysis models and image processing chains which leverage distributed storage and analytics**
  - Raster Analysis works with your existing GIS data and imagery
    - register your data with Image Server without converting
  - Raster Analysis can optimize your data for distributed analytics
    - result imagery is written into distributed raster storage for improved scalability
  - Raster Analysis is designed to scale with your organization’s demands
    - scale up to get the job done quicker, scale down when resources are not needed
Enterprise Image Server with Distributed Raster Analysis

- Complete analysis task for data too big for single desktop machine
  - Meet time constraint (Month -> Weeks -> Days -> Hours -> Minutes)

Dynamic Raster Models
Geoprocessing Models
Server-based Distributed Raster Analytics with Distributed Raster Data Storage
Web GIS Layers

on-the-fly processing
powerful analytics
(persistent) distributed analytics with optional distributed storage for even greater scalability
rich geoinformation model
**Raster Analysis Capabilities**

**Large Collection of Raster Functions**

**Multiband Math**

<table>
<thead>
<tr>
<th>Math</th>
<th><strong>Addition</strong></th>
<th><strong>Subtraction</strong></th>
<th><strong>Multiplication</strong></th>
<th><strong>Division</strong></th>
<th><strong>Exponential</strong></th>
<th><strong>Logarithmic</strong></th>
<th><strong>Trigonometric</strong></th>
<th><strong>Rounding</strong></th>
<th><strong>Miscellaneous</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Add</td>
<td>A + B</td>
<td>A - B</td>
<td>A * B</td>
<td>A / B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Sub</td>
<td>A - B</td>
<td>A - B</td>
<td>A * B</td>
<td>A / B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Mul</td>
<td>A * B</td>
<td>A * B</td>
<td>A * B</td>
<td>A / B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Div</td>
<td>A / B</td>
<td>A / B</td>
<td>A / B</td>
<td>A / B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Exp</td>
<td>A^B</td>
<td>A^B</td>
<td>A^B</td>
<td>A^B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Exp10</td>
<td>A^10</td>
<td>A^10</td>
<td>A^10</td>
<td>A^10</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Exp2</td>
<td>A^2</td>
<td>A^2</td>
<td>A^2</td>
<td>A^2</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Plus</td>
<td>A + B</td>
<td>A - B</td>
<td>A * B</td>
<td>A / B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
<tr>
<td>Power</td>
<td>A^B</td>
<td>A^B</td>
<td>A^B</td>
<td>A^B</td>
<td>A^B</td>
<td>Log(A)</td>
<td>Sin(A)</td>
<td>Round(A)</td>
<td>Negate</td>
</tr>
</tbody>
</table>

**Correction**

- Apparent Reflectance
- Geometric Correction
- Specular Filtering (Lee, Frost, Kuan)
- Thermal noise
- Radiometric Calibration

**Data Management & Conversion**

- Raster to Vector
- Vector to Raster
- Colormap
- Colormap To RGB
- Grayscale
- Remap / Reclass
- Spectral Conversion
- Unit Conversion
- Vector Field
- LAS to Raster
- LAS Dataset to Raster
- Clip
- Composite
- Extract Bands
- Mask
- Mosaic Rasters
- Rasterize Features
- Reproject
- Nibble

**Visualization & Appearance**

- Contrast and Brightness
- Convolution
- Pan sharpening
- Resample
- Statistics and Histogram
- Stretch

**Interpolation**

- Interpolate Irregular Data
- Nearest Neighbor
- IDW
- EEBK
- Swath

**Surface Generation & Analysis**

- Aspect
- Curvature
- Elevation Void Fill
- Hillshade
- Shaded Relief
- Slope
- Viewshed

**Analysis: Distance & Density**

- Euclidean Distance
- Cost Distance
- Least Cost Path
- Resample
- Statistics and Histogram
- Stretch

**Analysis: Overlay**

- Weighted Overlay
- Weighted Sum

**Analysis: Band Math & Indices**

- NDVI / NDVI Colorized
- SAVI / MSAVI / TSAVI
- GEMI
- GVI (Landsat TM)
- PVI
- Tasseled Cap (Kauth-Thomas)
- Binary Thresholding

**Analysis: Image Segmentation & Classification**

- Segmentation (Mean Shift)
- Training (ISO, SVM, ML)
- Supervised Classification

**Analysis: Hydrology**

- Fill
- Flow Accumulation
- Flow Direction
- Flow Distance
- Stream Link
- Watershed

---

Chain functions together into Raster Models and apply them to answer complex questions.
ArcGIS Python API
ArcGIS Python API overview

- The ArcGIS API for Python is implemented using the online and on-premises web GIS platform provided by ArcGIS Online and ArcGIS Enterprise respectively.
- Distributed as the `arcgis conda` package.
- `arcgis.raster` module
  - The `arcgis.raster` module contains classes and `raster analysis functions` for working with raster data and imagery layers in ArcGIS Enterprise.
ArcGIS Python API for Raster Analysis - Image Layer

- Display Image Layer
- Apply Image Service on-the-fly processing
ArcGIS Python API for Raster Analysis – Raster Function

- Construct custom raster function chain
- Preview analysis result
- Submit raster analysis task
ArcGIS Python API for Raster Analysis – Hydrology Workflow

- Global Analysis function generates persisted output before the next task.
Raster Analysis using Python API
Raster Analysis
System Deployment
Deployment Distributed Raster Analysis
Deployment platform

- **On-premise + network file shared storage appliance**
- **Cloud platform + Cloud Object Storage**
  - ArcGIS Cloud Builder for Azure
  - Amazon Cloud Formation template or Command line tool

Raster Analysis
System Components
Raster Store

- **New ArcGIS server data store type**
- **Raster Store registered to Image Server is for storing output imagery of Raster Analysis tasks**
  - Shared file system storage
  - Cloud Object Storage
    - Amazon S3
    - Azure Blob Storage

  **Note:** register a Cloud Store first then register a Raster Store reference to the Cloud Store

- **Feature output of Distributed Raster Analysis stored in Hosting server’s**
Optimized Cloud Raster Format (CRF)

- **New ArcGIS Raster Dataset format for distributed read/write**
  - Default output imagery format for Raster Analysis tasks
  - Image is split to “bundles” with optimized schema
  - Support cloud storage and file system
  - Published as “hosted” image service
  - Rendered through Raster Rendering service
Distributed Raster Analysis System Services

- Geo Processing service as master
- Image Service as worker
- Image Service for rendering output
Raster Analysis Tools System Service

Folder: System
Current Version: 10.6
View Footprints In: ArcGIS Online Map Viewer

Services:
- System/CachingControllers (GPServer)
- System/CachingToolsEx (GPServer)
- System/CachingTools (GPServer)
- System/DistributedWorker (GPServer)
- System/PublishingToolsEx (GPServer)
- System/PublishingTools (GPServer)
- System/RasterAnalysisTools (GPServer)
- System/RasterProcessing (ImageServer)
- System/RasterRendering (ImageServer)
- System/ReportingTools (GPServer)
- System/SceneCachingControllers (GPServer)
- System/SceneCachingTools (GPServer)
- System/SyncToolsEx (GPServer)
- System/SyncTools (GPServer)
- System/UtilityNetworkTools (GPServer)

System/RasterAnalysisTools (GPServer)
Service Description: The RasterAnalysisTools service is used by

Tasks:
- GenerateRaster
- ConvertFeatureToRaster
- CopyRaster
- SummarizeRasterWithin
- ConvertRasterToFeature
- CalculateDensity
- CreateViewshed
- InterpolatePoints
- Classify
- Segment
- TrainClassifier
- GenerateRasterCollection
- StreamLink
- Watershed
- FlowDirection
- FlowAccumulation
- Fill
- DetermineOptimumTravelCostNetwork
- DetermineTravelCostPathsToDestinations
- CalculateDistance
- FlowDistance
- Nibble
- CalculateTravelCost

Note: Generate Raster tool is the generic service tool takes raster function as input to support various analytic workflows

ref: http://esriurl.com/rasterfunc
Types of Distributed Raster Analysis Operations

- Different type of operation poses different challenge to parallelization
- Distributed Raster Analysis supports
  - Local/Focal analysis
    - Most raster functions
    - Custom python raster functions
  - Zonal analysis
    - Zonal statistics
  - Global analysis
    - Hydrology Analysis
    - Cost Distance

ref: http://esriurl.com/typeofops
Case Study: Hydro workflow scalability

- Computing hydrologic characteristics and flow patterns for US
  - For predicting stream flow and flood forecasting
- Old single thread tools required manual chunking of input
  - Performance suffered over half billion cells
- Now running 500x larger data on a distributed cluster

30m US - 25 Billion cells
57 separate processing units

10m US - 230 Billion cells
Enterprise Raster Analysis Clients

- ArcGIS Pro/Desktop
- Portal Webmap
- ArcGIS API for Python
Best practice, tips and tricks

• Split Raster Analytic and Image Hosting roles
• Increase server maximum heap memory for processing very large image
• Adjust maximum number of processing service instances based on machine specs
• A single raster analysis task is always default to use up to 80% of all the available raster processing service processes