High Performance Batch Geocoding

Bruce Harold  bharold@esri.com
Victor Bhattacharyya  vbhattacharyya@esri.com
Public and private organizations depend on geocoding
Addresses hard to control as a data type
Spatial enablement of data often comes late in the data lineage
ArcGIS supports Batch Geocoding in this situation
- Core geoprocessing tool with local or service based locators
- Publication of portal items
- Esri premium Apps (Insights for ArcGIS, Maps for Office)
- Custom Apps or geoprocessing tools calling the REST API
- Partner apps calling the REST API
Geocode Addresses
Core Geoprocessing

- Table Input
  - CSV, XLSX, GDB tables
- Locator
  - Local and Server locators
- Pro has better performance than ArcMap
  - Memory headroom
  - Can use > 2GB data cache
ArcGIS Online Batch Geocoding

Create Item Content

- Interactive Workflow
  - Add CSV Layer to Web Map
  - Limited to 1000 Features
  - Can be saved as a feature service

- Add Item & Publish Workflow
  - Upload CSV Item
  - Publish immediately or later
  - Unlimited Feature Count
  - ...Web Map displays 1000 only
  - ...ArcGIS Pro displays any number
ArcGIS Online Batch Geocoding
CSV to Layer
Maps for Office
Microsoft Excel

- Spreadsheet to Map
- Slides
- Analysis
New in 10.5.1 Enterprise
Geocode Locations from Table

Batch Geocoding in Esri Maps for Office

Batch Geocoding in Excel
How Does It Work In Pro?
Geocoding

- Pro ships with default locators
  - XY provider for coordinate and grid handling
  - Esri World Geocoder (requires internet)
- Locate pane defines available locators
  - Add more to your project
- Build your own locators
  - Locators built in Pro or Desktop 10.x work
- 64bit memory headroom improves performance
- Future release will support World service Category filtering
  - Specify the type of address or feature found
Your Data
Tabular Presentation Affects Quality & Performance

• Language doesn’t matter
  - *Esri Locators* understand multiple languages and transliteration
  - Build your locators in your language(s)
  - **Languages should not be mixed within a record - duplicate reference data per language**

• You don’t always have control over data schema or quality
  - You can use any supported tabular format: CSV, Excel, dBase, Database…
  - Varying input fields can frustrate downstream Geoprocessing
  - Consider ETL work to get your data into consistent shape, it will live for a long time

• If your data allows, map fields to the Multiple Fields locator inputs
• If not, *make* a SingleLine field
  - AddressLine1, 2, 3 etc. concatenated with space separator
  - In Geoprocessing terms: Table to Table
Your Data

**Singleline**

- Making a Singleline Field
  - You don’t need a fancy ETL workflow
  - Use the Table to Table Geoprocessing tool
Geocode Addresses

Core Geoprocessing

- **Table Input**
  - Views, Selections *not* honored
  - Watch this when using ArcGIS Online
- **Runtime Properties Are Used**
  - Data Cache Memory
  - Thread Count
  - Match Score Threshold
  - Geometry from XY Display Fields
  - ...
- **Pro has better performance**
  - Supports larger data cache than 10.x (>2GB)
Core Geoprocessing Recommendations #1
Batch Geocoding With Local Locators

• Use File Folder storage for locators
  - ArcGIS 10.4 deprecated Geodatabase storage
  - ArcGIS Pro *enforces* file folder storage

• Use Solid State Disk for Locators

• Specify Data Cache Size as large as practicable
  - Locator data is decompressed into RAM to this limit
  - Pro and Background Geoprocessing can use enough RAM to hold an entire locator
  - OR…

• Esri can supply completely uncompressed StreetMap locators on request
  - Faster startup, faster processing
Core Geoprocessing Recommendations #2

Leverage Threads

- File-based locators can leverage available cores by means of threading
- Set Number of Threads to Auto for single locators
  - System will use one less thread than it sees CPU cores
- Composite locators can use multiple threads too
  - No Auto option, threading properties of member locators are ignored
  - Has to be enabled in the .LOC file that defines the composite
  - Spread the thread count amongst member locators by expected workload
Recommendation #3 - Sort Strategy
Use the Locator Batch Presort Property

• Locators are spatially organized and accessed
  - Descending spatial extents (State, City, ZIP)
• Batches are chunked behind the scenes
• If you mix zones per chunk it is expensive
• Sort your data by all zone fields
How should Geocode Addresses Perform?

**Tuned Data and Locators**

- USA Composite, Tuned, 2GB Cache
- SSD, 8 Cores, 16GB RAM
- 1 Million USA Nationwide Rows
- Sorted Table, Multiple Fields
- Performance OK but 10.x decays 40%
  - Started out at 1.5M/hr
  - Pro better
- What is a good speedup strategy?
  - Embarrassingly parallel problem
Core Demo
Locator Properties
Geocode Addresses
Custom Geoprocessing

Splitting jobs or the machine
Splitting Your Machine

Chunk Geocoding Using Python’s Subprocess Module

- Master Script Tool
- Worker Script
- Is this a good idea?
- Demo and Discussion

Script tool creates workspaces, splits the job, runs concurrent geocoding and merges results.
Splitting Jobs
Concurrency

• Multiprocessing Examples
  - Data Interoperability & ArcGIS Online
  - Custom Script Tool & Enterprise
Truly Parallelized Batch Geocoding

Client-Server Architecture

- ArcGIS Pro custom script tools
  - Python module concurrent.futures
- Leverage Geocode Server instances
  - REST API call `geocodeAddresses`
  - Server or Portal
  - StreetMap Premium or World Geocoder for ArcGIS
- Input is read in zone field sort order
- Chunks of 1000 rows sent
- Number of concurrent chunks = instance count
- Chunk processing interleaved
- Scales linearly with service instance count
Custom GP Demo

Subprocesses
Data Interoperability
Custom Script Tool
Geocode Locations from Table
New at Enterprise 10.5.1
Geocode Locations from Table
On-Premises Map Viewer

- A new Map Viewer Analysis tool for geocoding large tables located on your Portal
  - Recommended way to geocode large tables quickly
  - Returns all output fields from the geocoding service

- Works using geocoding services that are federated with your portal
  - Including Esri’s World Geocoding Service and Locator Services hosted by your organization

- Allows your administrator to optimize geocoding
  - Administrators can configure the system to deliver performance and manage load
Geocode Locations from Table

Large Batch Geocoding in ArcGIS Enterprise

- Supports multiple input formats
  - CSV
  - XLSX
  - Portal table

- Geocoding jobs run asynchronously in the background

- Supports multiple output formats and writes geocoding results to the portal
  - CSV
  - XLS
  - Feature Layer
Portal Geocoding

Geocode Locations from Table in action
Configure “Geocode Locations from Table”

Geocoding Service Instances

Geoprocessing Service

GeocodeAddresses REST Call

Documentation
Configure “Geocode Locations from Table”

• How to Publish locators and add them as Utility Services

• Set the number of threads per geocoder via the Sharing API

• Scale the GP Service based on server resources via Server Manager

• Learn more at the Configure Portal to Geocode Addresses online help
Portal Demo
Configuring Parallelized Batch Geocoding

Configure Threads

(Elapsed Time: 28.73 seconds)

(Elapsed Time: 2 minutes 21 seconds)
Future Work
Large Batch Geocoding in ArcGIS Pro

• Adding a Pane in a future release of Pro
• Supports geocoding Portal tables using Portal locators
• Will support categories and source countries
Please Take Our Survey on the **Esri Events App**!

**Download the Esri Events app and find your event**

**Select the session you attended**

**Scroll down to find the survey**

**Complete Answers and Select “Submit”**
Questions?