

High Performance Batch Geocoding

Bruce Harold

bharold@esri.com

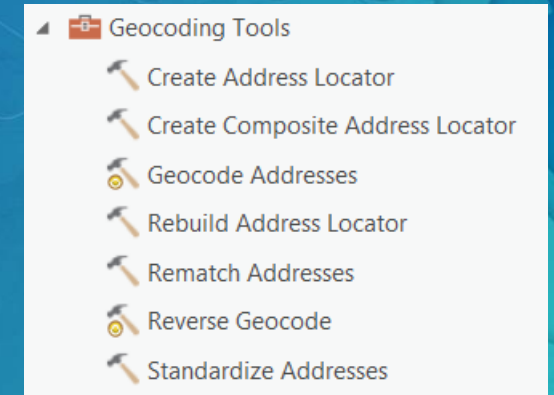
Victor Bhattacharyya

vbhattacharyya@esri.com

Batch Geocoding

Finding Locations From Descriptions

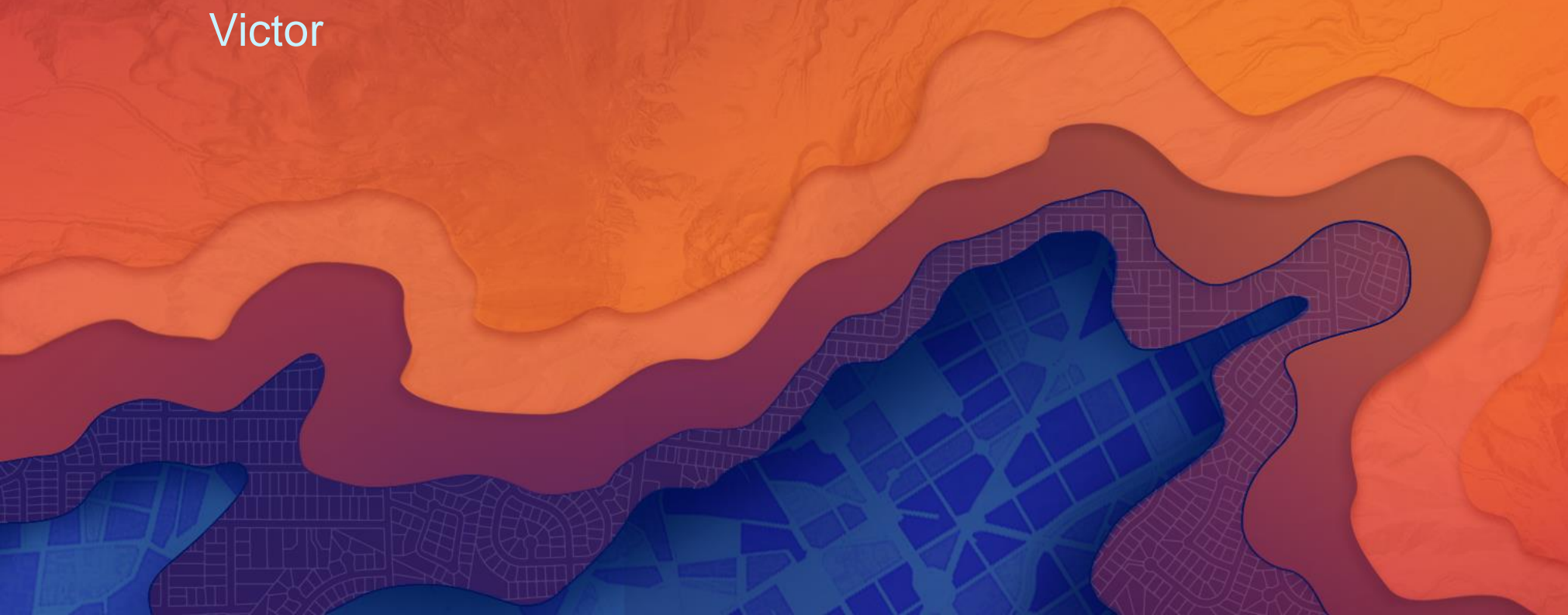
- Public and private organizations depend on geocoding
- Addresses hard to control as a data type
- Spatial enablement of data often comes late in the data lineage
- ArcGIS supports Batch Geocoding in this situation
 - Core geoprocessing tool with local or service based locators
 - Publication of portal 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

A screenshot of the 'Publish' dialog box in ArcGIS. The dialog box is titled 'Publish' and has a close button (X) in the top right corner. It contains the following fields and options:

- Title:** A text box containing 'TestData'.
- Tags:** A text box containing 'Batch' with a close icon (x). Below it is a link 'Add tag(s)'.
- Locate features using:** Three radio buttons: 'Latitude/Longitude' (unselected), 'Address' (selected), and 'None, add as table' (unselected).
- Country:** A dropdown menu showing 'United States'.
- Review the field types and location fields. Click on a cell to change it.** Below this is a table with three columns: 'Field Name', 'Field Type', and 'Location Fields'.

Ways to Batch Geocode

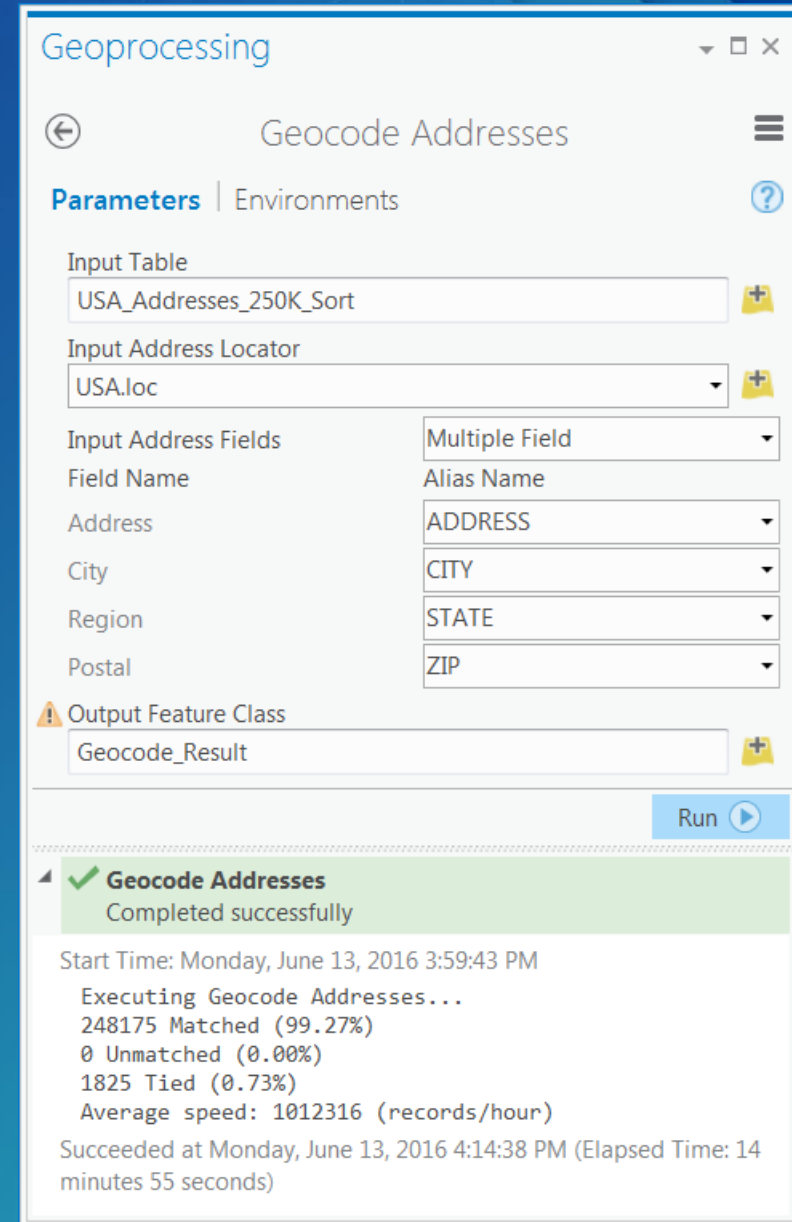
Victor



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



Geoprocessing

Geocode Addresses

Parameters | Environments

Input Table
USA_Addresses_250K_Sort

Input Address Locator
USA.loc

Input Address Fields	Multiple Field
Field Name	Alias Name
Address	ADDRESS
City	CITY
Region	STATE
Postal	ZIP

Output Feature Class
Geocode_Result

Run

✓ **Geocode Addresses**
Completed successfully

Start Time: Monday, June 13, 2016 3:59:43 PM
Executing Geocode Addresses...
248175 Matched (99.27%)
0 Unmatched (0.00%)
1825 Tied (0.73%)
Average speed: 1012316 (records/hour)
Succeeded at Monday, June 13, 2016 4:14:38 PM (Elapsed Time: 14 minutes 55 seconds)

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

Add CSV Layer ✕

Locate features using: ☐ Latitude/Longitude ☒ Address

Country:

Review the location fields

Field Name
OBJECTID
Singleline

Item from my computer ?

Add an item from your computer.

File: TestData.csv

Title:

Tags:

☒ Publish this file as a hosted layer
(Adds a hosted layer item with the same name.)

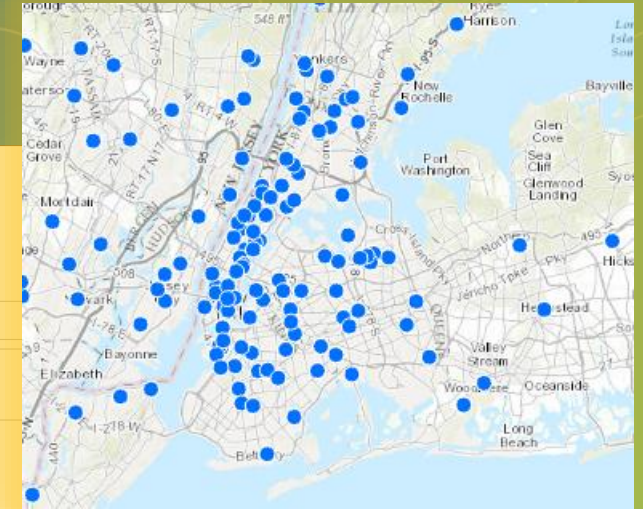
Use: ☐ Latitude/Longitude ☒ Address ☐ Table

Country:

Review the field types and location fields. Click on a cell to change it.

Field Name	Field Type	Location Fields
ADDRESS	String	Street
CITY	String	City
STATE	String	State
ZIP	String	ZIP

Time Zone: ?



Add CSV Layer

Locate features using: ☐ Latitude/Longitude ☒ Address

Country:

Review the location fields

Field Name
OBJECTID
Singleline

Item from my computer ?

Add an item from your computer.

File: TestData.csv

Title:

Tags:

☒ Publish this file as a hosted layer
(Adds a hosted layer item with the same name.)

Use: ☐ Latitude/Longitude ☒ Address ☐ Table

Country:

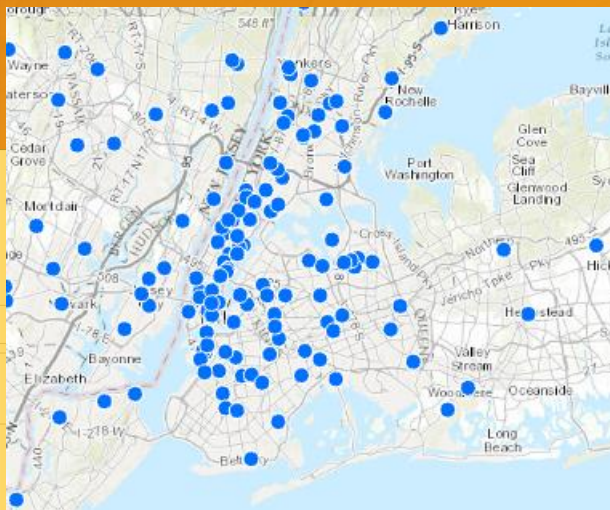
Review the field types and location fields. Click on a cell to change it.

Field Name	Field Type	Location Fields
ADDRESS	String	Street
CITY	String	City
STATE	String	State
ZIP	String	ZIP

Time Zone:

ADD ITEM

CANCEL



ArcGIS Online Batch Geocoding

CSV to Layer

Microsoft Excel

- Spreadsheet to Map
- Slides
- Analysis

Add data from my spreadsheet

Verify the recommended mapping option

1 Data 2 Location

\$B\$1:\$E\$4001

Address

Confirm location

Types (Unique symbols)

Selected

Heat Map

Select

Credit usage: 160
Credits available: 64,651

☐ Don't show me confirm options again.

☐ Rotate symbols

Transparency

Overall

160 service credits of 64,651 available

Add data Cancel

Layer style
Change the appearance of your layer

\$B\$1:\$E\$4001 by STATE

Choose another style

STATE

Click to edit symbol or label.

	LABEL	COUNT
⋮	CA	358
⋮	TX	318
⋮	FL	260
⋮	NY	217
⋮	PA	178
⋮	IL	175
⋮	OH	164
⋮	NC	158

☐ Rotate symbols (degrees)

Transparency

Overall

OK Cancel

Batch Geocoding in Esri Maps for Office

Batch Geocoding in Excel

Add data from my spreadsheet

Verify the recommended mapping options.

1 Data 2 Location Type

\$B\$1:\$E\$4001 Address

Confirm location columns

Types (Unique symbols) ☒

Heat Map ☐

160 service credits of 64,651

Layer style

Change the appearance of your layer

\$B\$1:\$E\$4001 by STATE

Choose another style

STATE

Click to edit symbol or label.

LABEL	COUNT
CA	358
TX	318
FL	260
NY	217
PA	178
IL	175
OH	164
NC	158

☐ Rotate symbols (degrees)

Transparency

Overall

OK Cancel

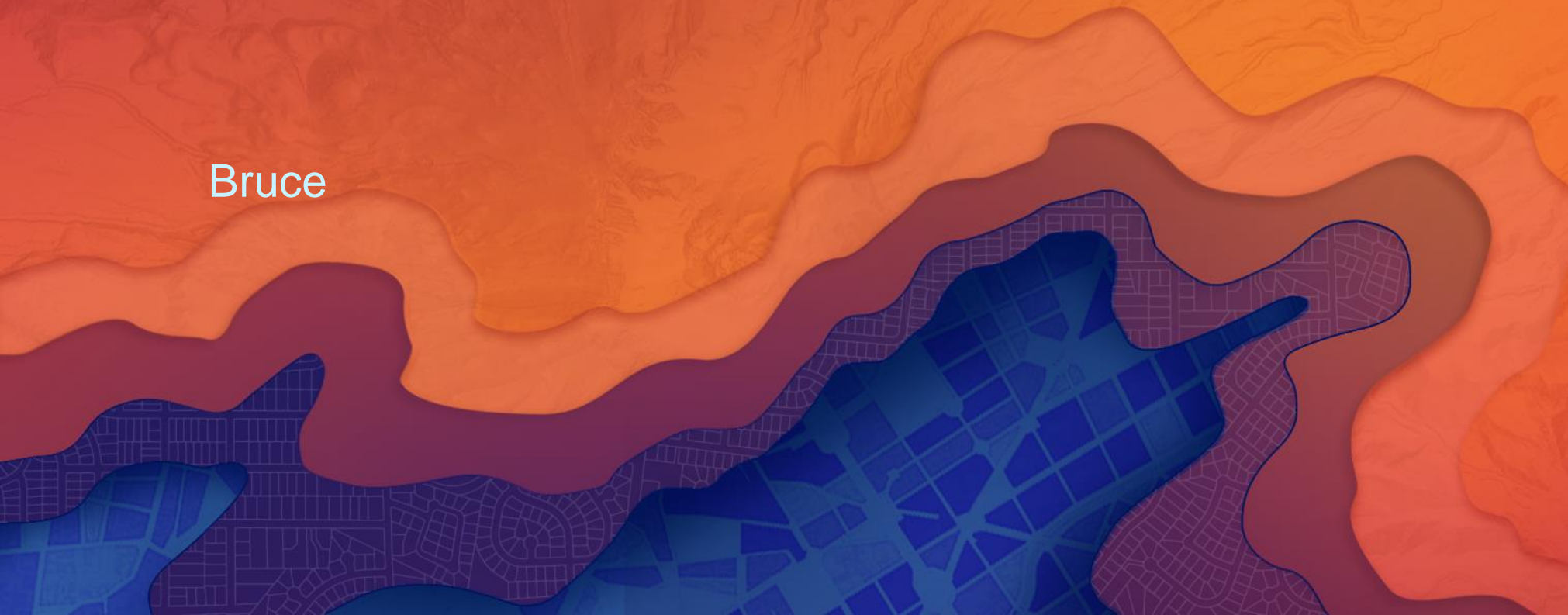
☐ Don't show me confirm options again.

Add data Cancel

Core Geoprocessing

Geocode Addresses

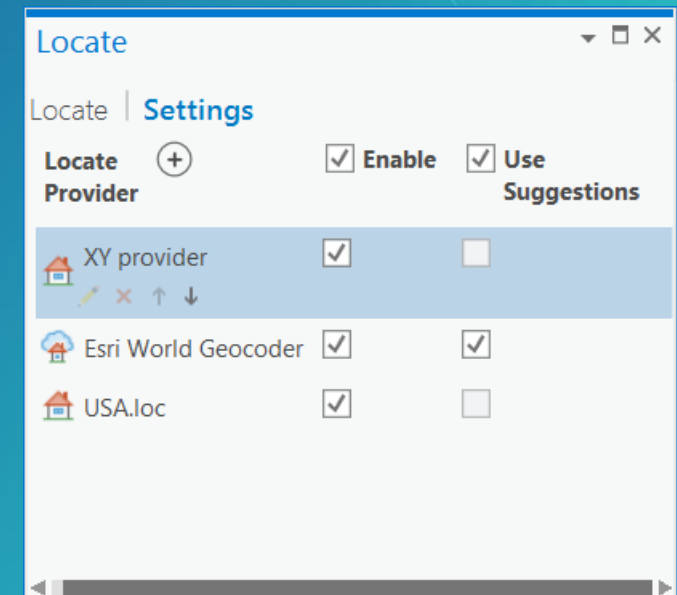
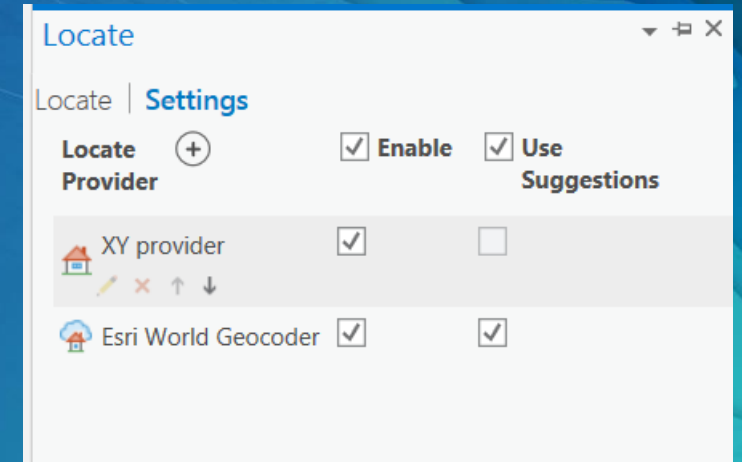
Bruce



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

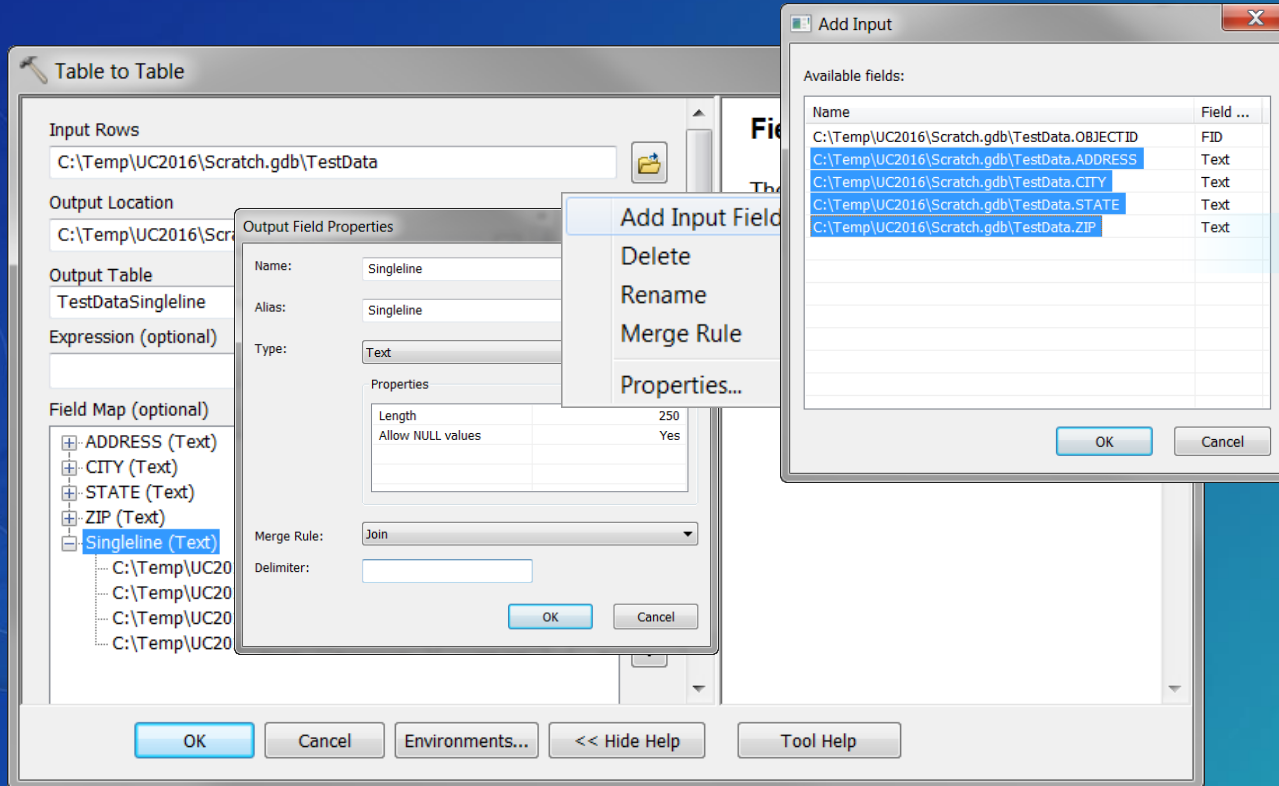
Tabular Presentation Affects Quality & Performance

- 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

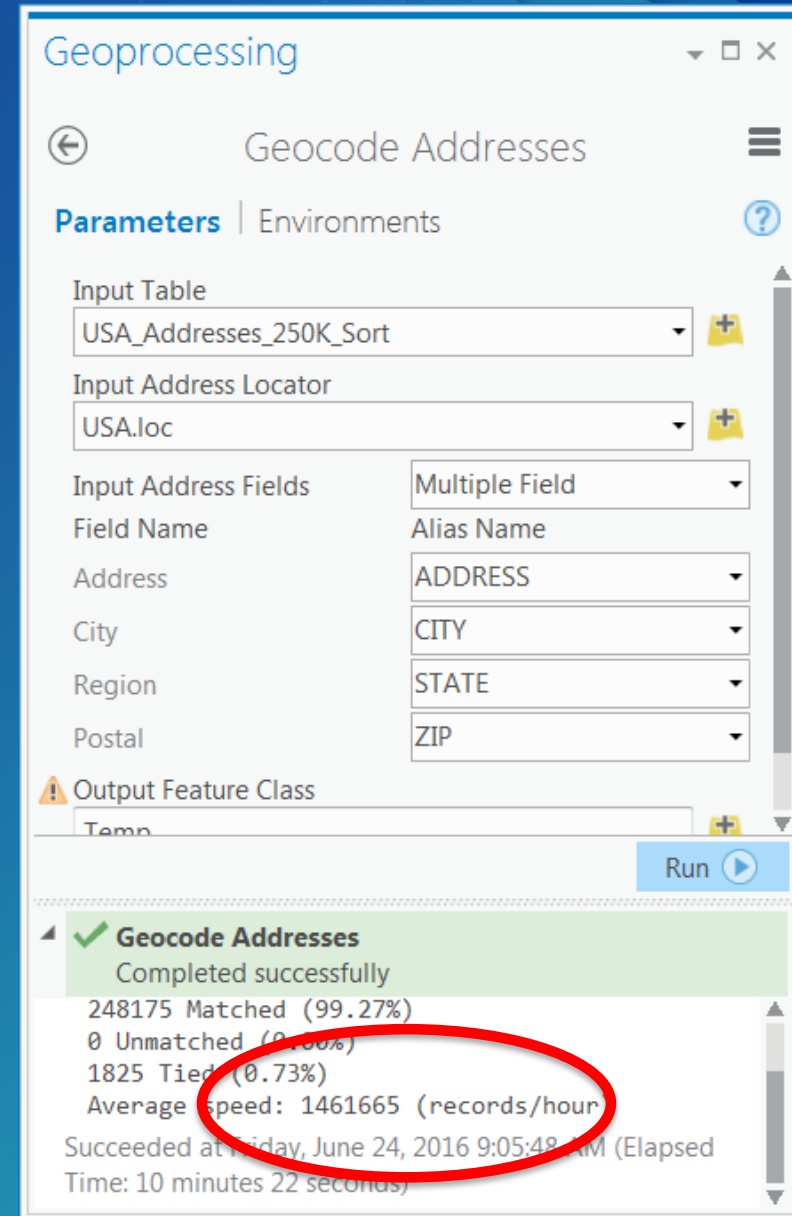


Singleline
7801 NW COUNTY ROAD 235 ALACHUA FL 32615
1550 W BERRY ST FORT WORTH TX 76110
2400 BYBERRY RD BENSLEM PA 19020
5401 GLENVIEW LN SAINT CLOUD MN 56303
700 MARSHFIELD AVE BERTRAND NE 68927
20900 LAWRENCE 1166 VERONA MO 65769
4300 S RIVER RD WEST SACRAMENTO CA 95691
201 14TH ST W GLENCOE MN 55336
1201 S UNION AVE CHICAGO IL 60607
4000 W 161ST ST CLEVELAND OH 44135
2 SYCAMORE ST NEWPORT RI 02840
501 E LARCH ST LIBBY MT 59923
2 W SPIRIT AVE TOMAHAWK WI 54487
6330 RUSTICATED STONE AVE HENDERSON NV 89011
100 WESTBROOK DR OSHKOSH WI 54904
2500 COUNTY ROAD C RHINELANDER WI 54501
400 S LOCUST ST BUFFALO MO 65622
35600 GROVEWOOD DR EASTLAKE OH 44095
1100 COUNTY ROAD 333 N HENDERSON TX 75652
3166 WATERGATE PL INDIANAPOLIS IN 46224
20500 9TH AVE E SPANAWAY WA 98387
700 MURDOCK RD NE FORT PAYNE AL 35967
904 N 143RD CIR OMAHA NE 68154
134 CUSTER ST STAMFORD CT 06902
5901 EIBECK LN WEST HARRISON IN 47060
100 S MARTIN L KING BLVD LAS VEGAS NV 89106
2700 E ADOBE ST MESA AZ 85213
1101 ALMAC DR DULUTH MN 55810
3001 ROXBORO RD EULESS TX 76039
6308 WOODMAN AVE VAN NUYS CA 91401

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

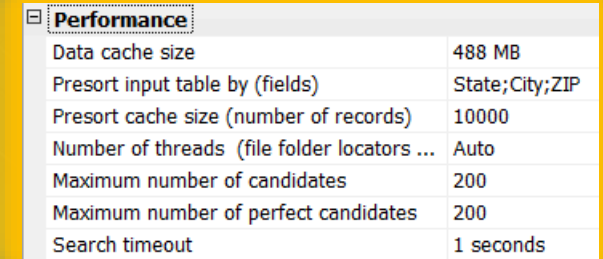
Batch Geocoding With Local Locators

- 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



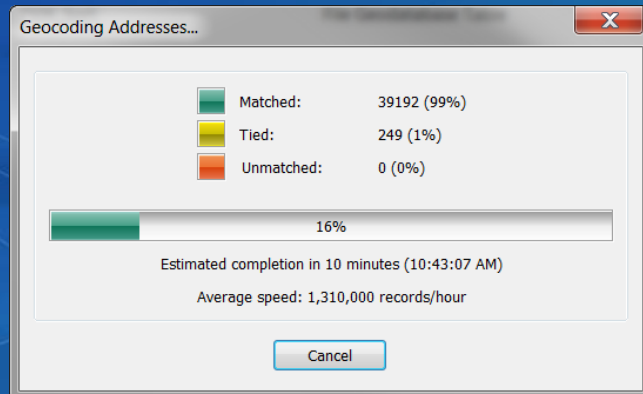
Performance	
Data cache size	488 MB
Presort input table by (fields)	State;City;ZIP
Presort cache size (number of records)	10000
Number of threads (file folder locators ...)	Auto
Maximum number of candidates	200
Maximum number of perfect candidates	200
Search timeout	1 seconds

- 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



Performance	
Data cache size	488 MB
Presort input table by (fields)	State;City;ZIP
Presort cache size (number of records)	2500

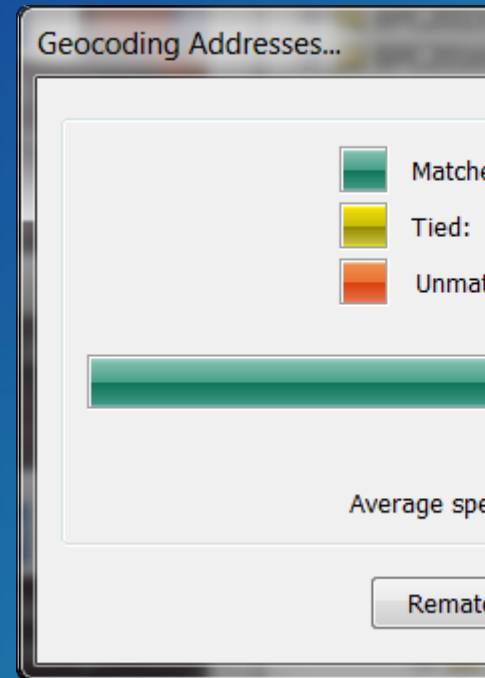


OBJECTID *	ADDRESS	CITY	STATE	ZIP
287	4445 ALSACE LOOP	FORT WAINWRIGHT	AK	99703
288	4055 TAMARACK DR	FORT WAINWRIGHT	AK	99703
289	4142 6TH ST	FORT WAINWRIGHT	AK	99703
290	2900 HC 60	HAINES	AK	99827
291	250 HERNDON DR	HOMER	AK	99603
292	38500 TRANQUILITY RD	HOMER	AK	99603
293	51500 SPAHR AVE	HOMER	AK	99603
294	55400 PREVET CT	HOMER	AK	99603
295	4161 PO BOX	HOMER	AK	99603
296	3684 MAIN ST	HOMER	AK	99603
297	2200 URSULA AVE	HOMER	AK	99603
298	200 STERLING HWY	HOMER	AK	99603
299	51000 TRAIL MOUNTAIN RD	HOMER	AK	99603
300	4047 MAIN ST	HOMER	AK	99603

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



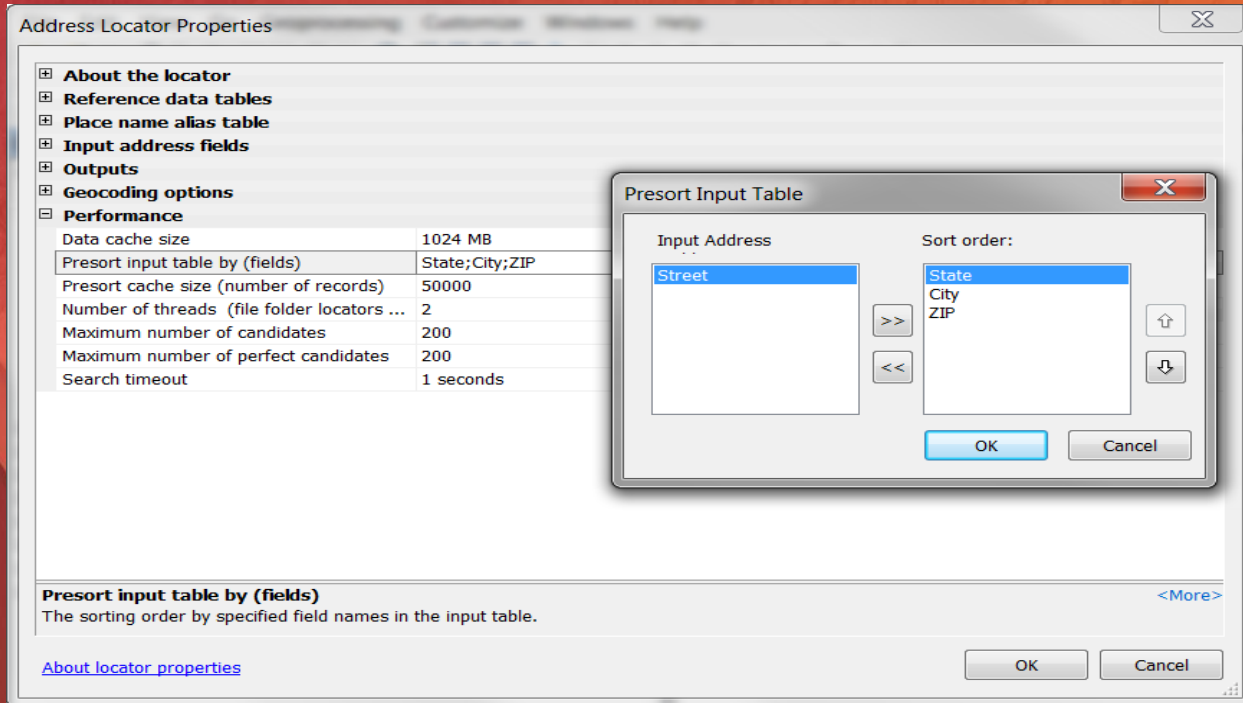
The 'Geoprocessing' window shows the 'Geocode Addresses' tool. The 'Parameters' tab is active, displaying the following settings:

- Input Table: USA_Addresses_1M_Sort
- Input Address Locator: USA.loc
- Input Address Fields: Multiple Field
- Field Name: Alias Name
- Street or Intersection: ADDRESS
- City or Placename: CITY
- State: STATE
- ZIP Code: ZIP
- Output Feature Class: USA1M_Result
- Dynamic Output Feature Class: ☐

The 'Run' button is located at the bottom right of the parameters section. Below the parameters, the execution results are displayed:

Geocode Addresses
Completed successfully

Executing Geocode Addresses...
936808 Matched (93.68%)
52069 Unmatched (5.21%)
11123 Tied (1.11%)
Average speed: 933892 (records/hour)
Succeeded at Friday, July 07, 2017 11:44:31 AM (Elapsed Time: 1 hours 4 minutes 35 seconds)

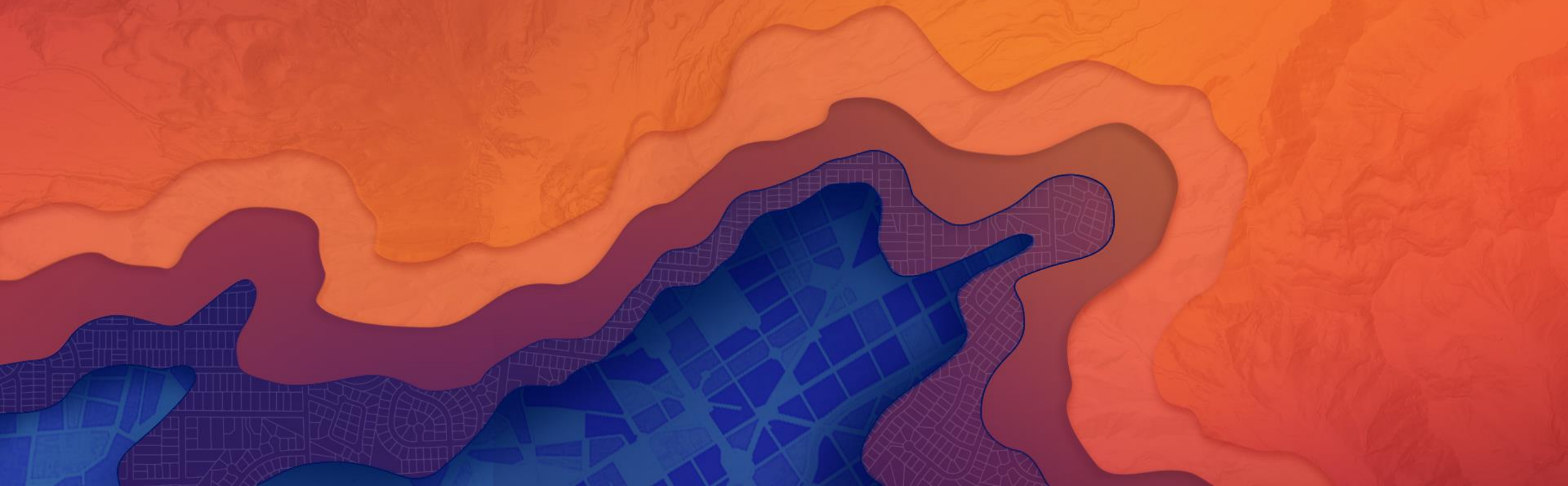


Core Demo

Locator Properties
Geocode Addresses

Custom Geoprocessing

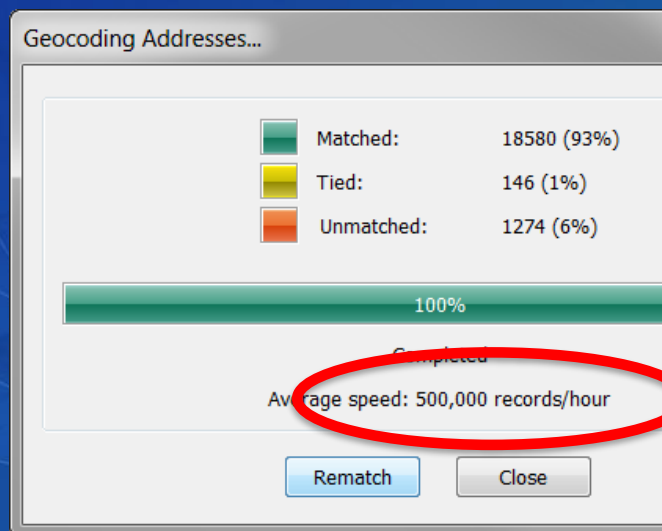
Splitting jobs or the machine



Splitting Your Machine

Chunk Geocoding Using Python's Subprocess Module

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



SubprocessGeocode

Completed

☐ Close this dialog when completed successfully

```
"MOD(OBJECTID,4) = 2" State;City;ZIP C:\Temp\UC2016\Locators\USA_StreetAddress "Street Address"
VISIBLE NONE;City City VISIBLE NONE;State State VISIBLE NONE;ZIP ZIP VISIBLE NONE"

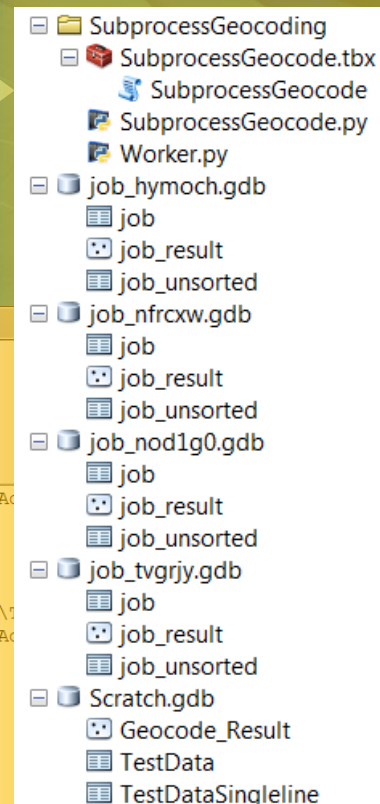
Starting process with command: C:\Python27\ArcGIS10.4\pythonw.exe C:\Temp\UC2016\SubprocessGeocoding\Worker.py C:\Temp\UC2016\job_nfrxw.gdb C:\Temp\UC2016\Scratch.gdb
"MOD(OBJECTID,4) = 3" State;City;ZIP C:\Temp\UC2016\Locators\USA_StreetAddress "Street Address"
VISIBLE NONE;City City VISIBLE NONE;State State VISIBLE NONE;ZIP ZIP VISIBLE NONE"

Waiting for geocode commencement: Wed Jun 15 10:31:22 2016

Waiting for geocode completion... Wed Jun 15 10:31:24 2016
Geocoded 4000 records from 20000 input...
Geocoded 12000 records from 20000 input...
Geocoded 20000 records from 20000 input...

Merging results...

Geocoded 20000 records at a net 771373 records per hour
Completed script SubprocessGeocode...
Succeeded at Wed Jun 15 10:32:33 2016 (Total Time: 1 minutes 42 seconds)
```

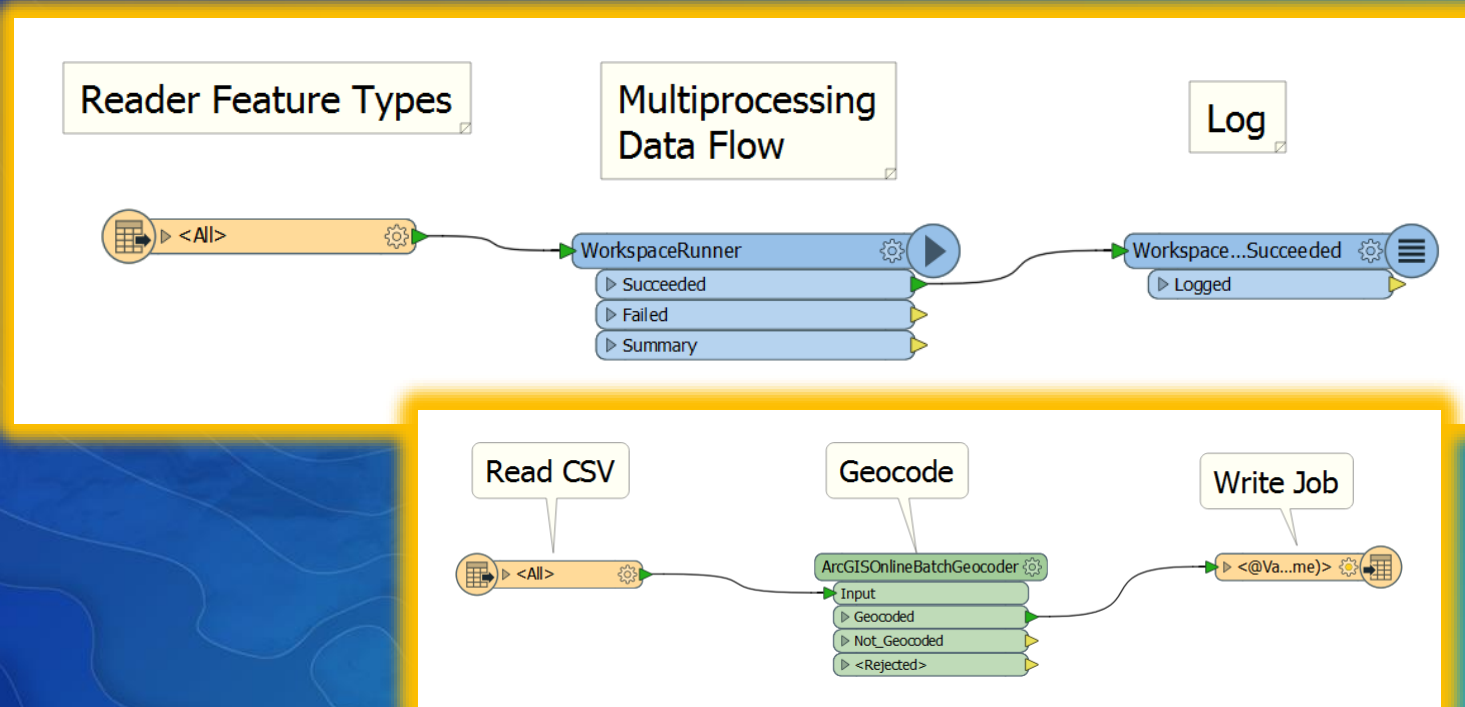


Script tool creates workspaces, splits the job, runs concurrent geocoding and merges results

Splitting Jobs

Concurrency

- Multiprocessing Examples
 - Data Interoperability & ArcGIS Online
 - Custom Script Tool & Enterprise



Geoprocessing

BatchSMPGeocoder

Parameters | Environments

GeocodeService
http://boris.esri.com:6080/arcgis/rest/services/USA/Geocode!

Input Table
TestData

Concurrency
8

Batch Mode
Multiple Fields

Address Field
ADDRESS

Neighborhood Field

City Field
CITY

Subregion Field

Region Field
STATE

Postal Field
ZIP

PostalExt Field

Category Filter
Address

Output Spatial Reference
4326

Output Feature Class
TestData_BatchSMPGeocoder

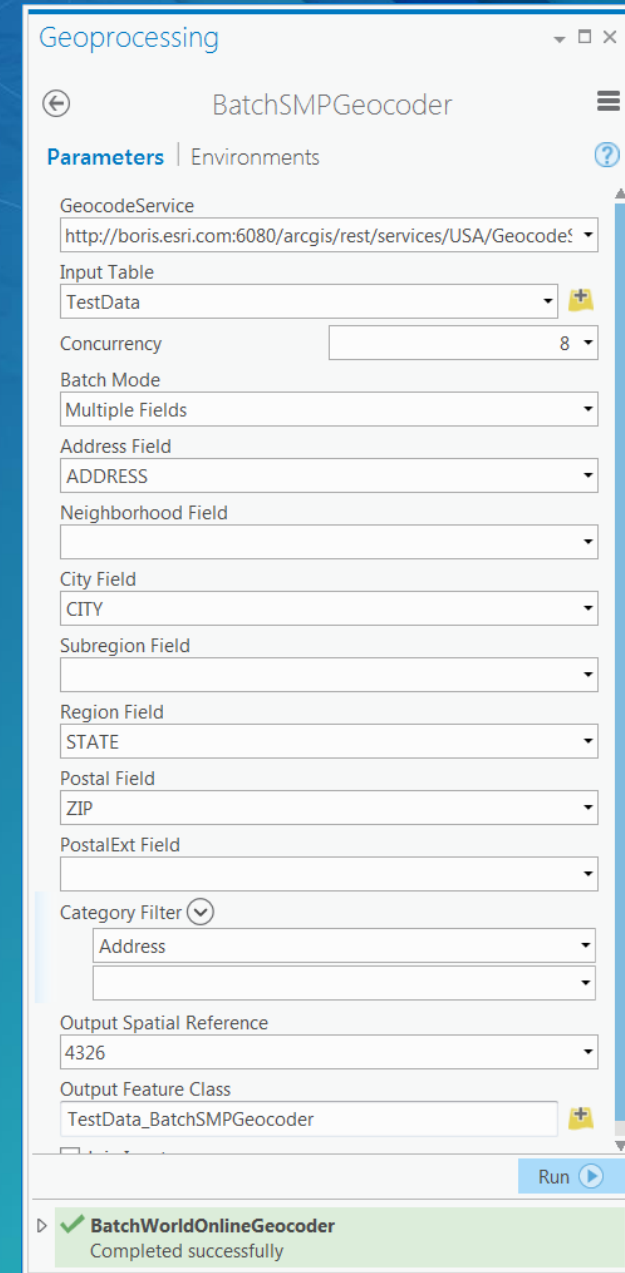
Run

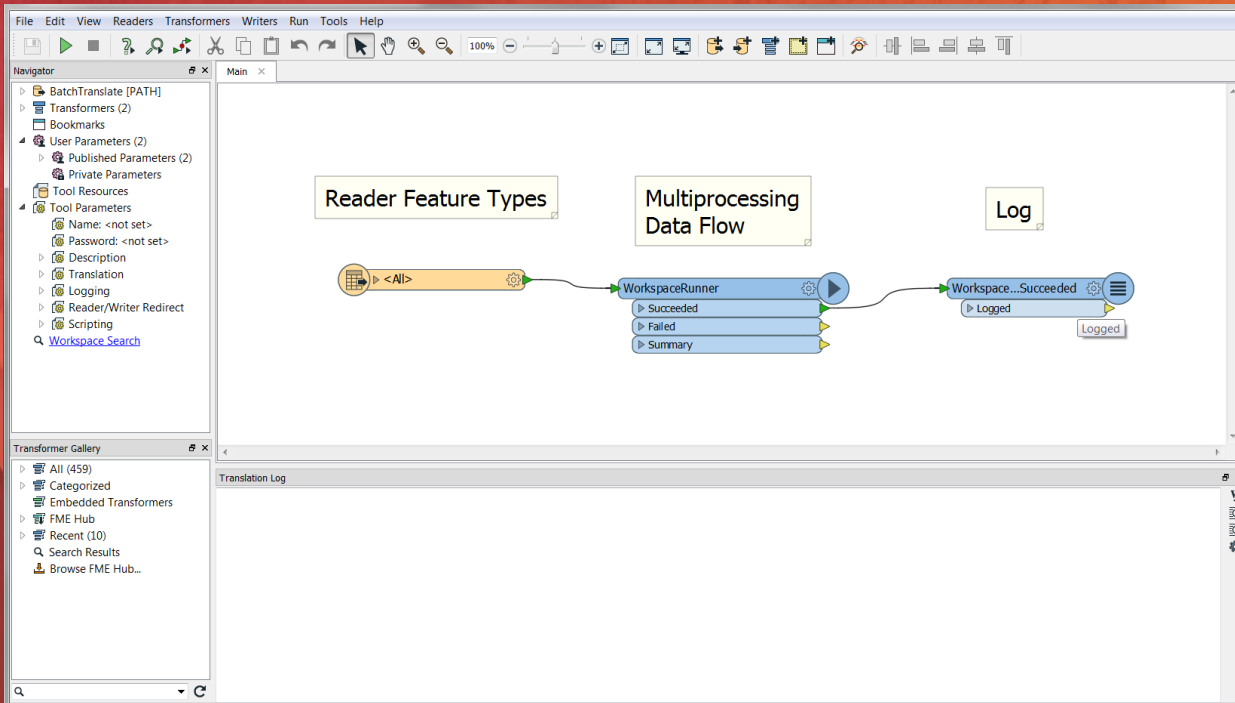
BatchWorldOnlineGeocoder
Completed successfully

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



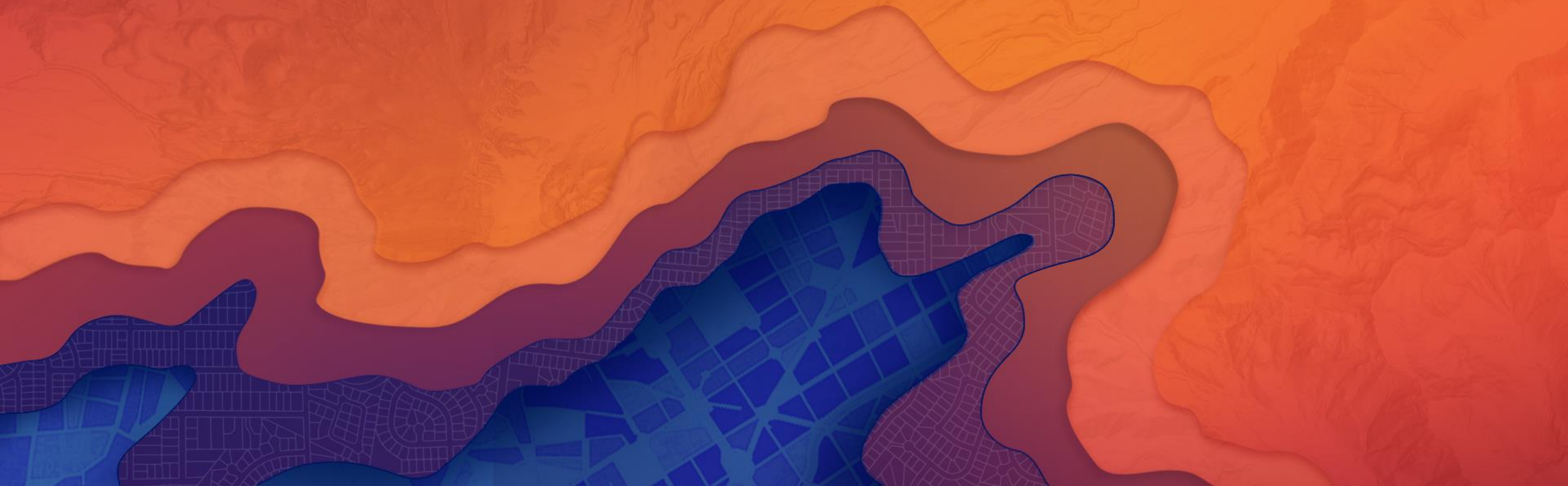


Custom GP Demo

Subprocesses
Data Interoperability
Custom Script Tool

Geocode Locations from Table

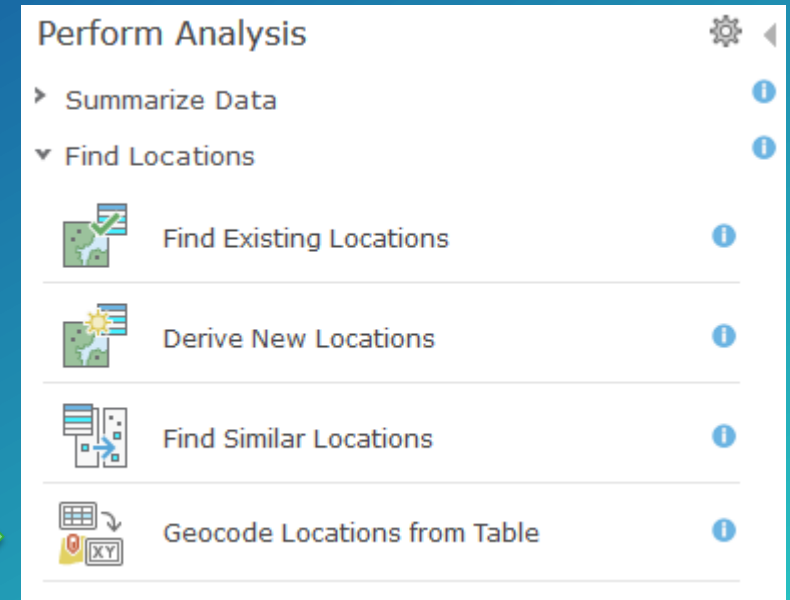
New at Enterprise 10.5.1



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

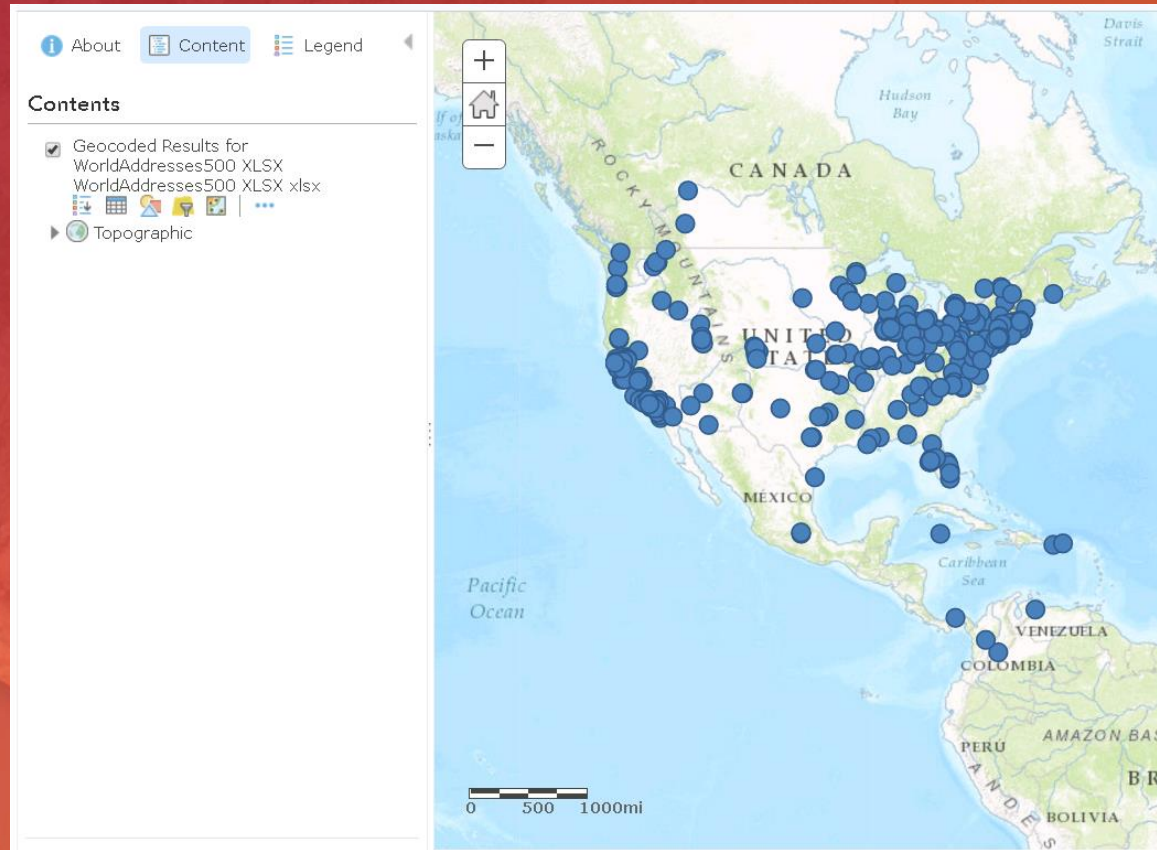
- Supports multiple input formats
 - CSV
 - XLSX
 - Portal table
- Geocoding jobs run asynchronously in the background
- Supports multiple output formats and writes geocoding results to the portal
 - CSV
 - XLS
 - Feature Layer

The screenshot shows the 'Geocode Locations from Table' tool interface. It is divided into four numbered steps:

- 1 Choose an input table**: A dropdown menu shows 'Address380-Address380'.
- 2 Choose a locator**: A dropdown menu shows 'Esri World Batch Geocoder'. Below it, a 'Country' dropdown shows 'World'. Under 'Select Data Fields', there are two buttons: 'Single Field' and 'Multiple Fields' (which is selected). A table below shows the mapping of locator inputs to data fields:

Locator Inputs	Data Fields
Address	address
Neighborhood	Not Used
City	city
Subregion	Not Used
Region	state
- 3 Choose an output format**: A dropdown menu shows 'CSV'.
- 4 Result layer name**: A text box contains 'Geocoded Results for Address380-Address380'. Below it, a 'Save result in' dropdown shows 'publisher'.

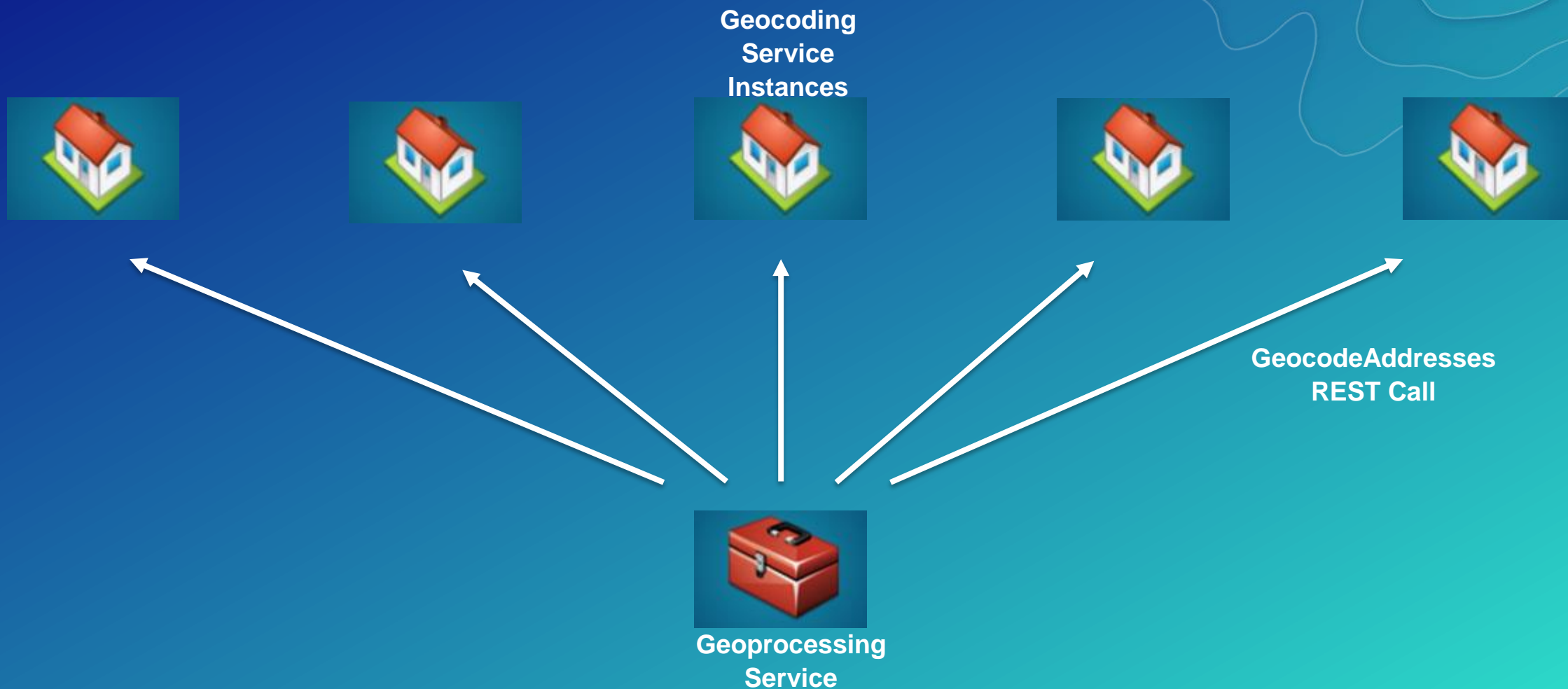
At the bottom of the interface is a large blue button labeled 'RUN ANALYSIS'.



Portal Geocoding

Geocode Locations from
Table in action

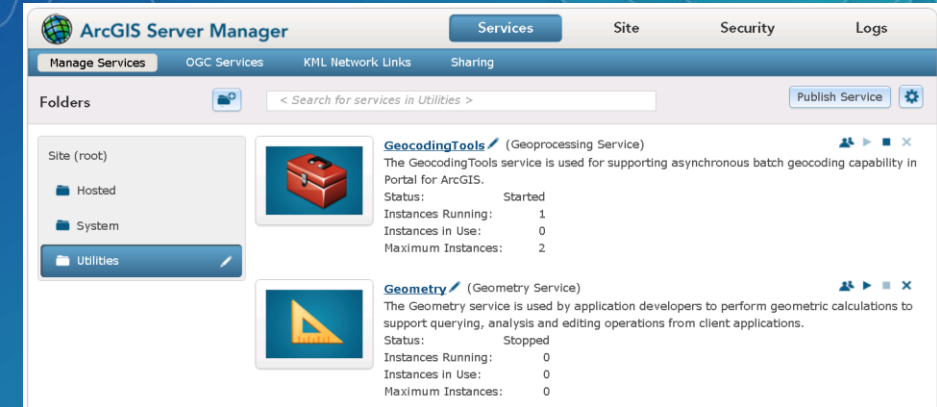
Configure “Geocode Locations from Table”



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



```
[{
  "url" : "https://geocode.arcgis.com/arcgis/rest/services/World/GeocodeServer",
  "northLat" : "Ymax",
  "southLat" : "Ymin",
  "eastLon" : "Xmax",
  "westLon" : "Xmin",
  "name" : "Esri World Geocoder",
  "suggest" : true
}, {
  "url" : "https://machine.domain.com/server/rest/services/Locators/USA/GeocodeServer",
  "name" : "Streetmap Premium USA Geocoder",
  "placeholder" : "Find address or place",
  "singleLineFieldName" : "SingleLine",
  "batch" : true,
  "placefinding" : true,
  "suggest" : true,
  "zoomScale" : 10000,
  "numBatchThreads" : 4
}]
```

Configure Threads

Documentation

```
[{  
  "url" : "https://geocode.arcgis.com/arcgis/rest/services/World/GeocodeServer",  
  "northLat" : "Ymax",  
  "southLat" : "Ymin",  
  "eastLon" : "Xmax",  
  "westLon" : "Xmin",  
  "name" : "Esri World Geocoder",  
  "suggest" : true  
}, {  
  "url" : "https://machine.domain.com/server/rest/services/Locators/USA/GeocodeServer",  
  "name" : "Streetmap Premium USA Geocoder",  
  "placeholder" : "Find address or place",  
  "singleLineFieldName" : "SingleLine",  
  "batch" : true,  
  "placefinding" : true,  
  "suggest" : true,  
  "zoomScale" : 10000,  
  "numBatchThreads" : 4  
}]
```

Configure Threads

(Elapsed Time: 28.73 seconds)

(Elapsed Time: 2 minutes 21 seconds)

Portal Demo

Configuring Parallelized Batch
Geocoding

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

The screenshot shows the 'Geoprocessing' window with the tool 'Geocode Locations from Table' selected. The interface includes tabs for 'Parameters' and 'Environments'. The 'Parameters' tab is active, showing the following fields:

- Input Table:** A text box containing 'customers.csv' with a file selection icon to its right.
- Input Locator:** A dropdown menu currently set to 'Atlanta'.
- Field Mapping:** A section with two columns: 'Locator Inputs' and 'Data Inputs'.

Locator Inputs	Data Inputs
Street	Address
City	Not Used
State	Not Used
Zip	Postal
- Output Name:** A text box containing 'Atlanta_geocoded'.
- Country:** A dropdown menu.
- Category:** A dropdown menu.

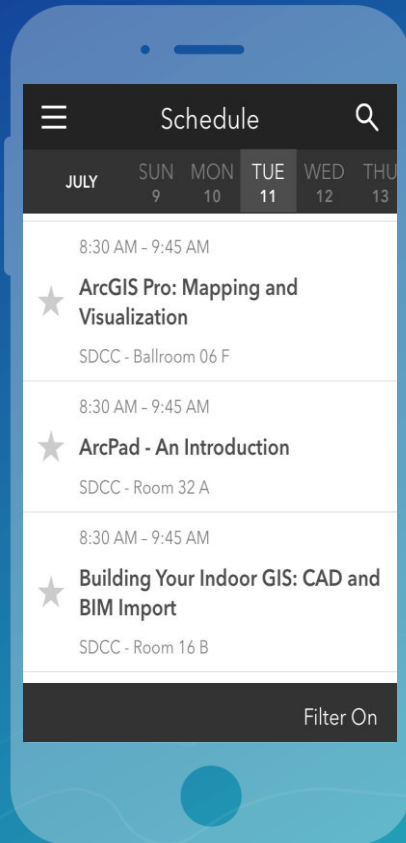
A 'Run' button with a play icon is located at the bottom right of the tool window.

Please Take Our Survey on the Esri Events App!

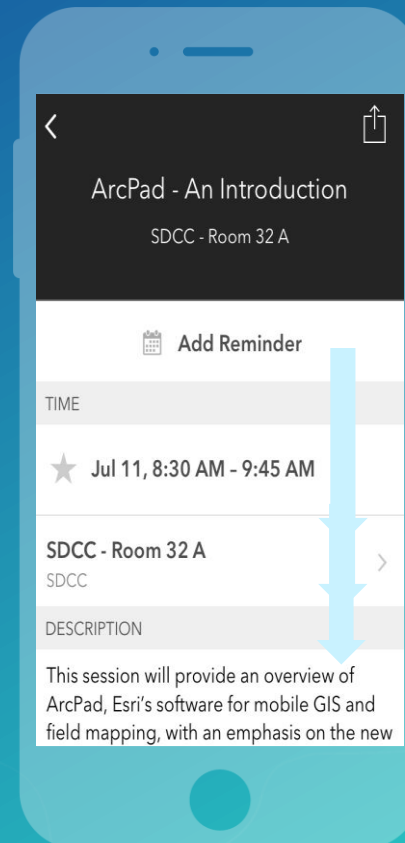
Download the Esri Events app and find your event



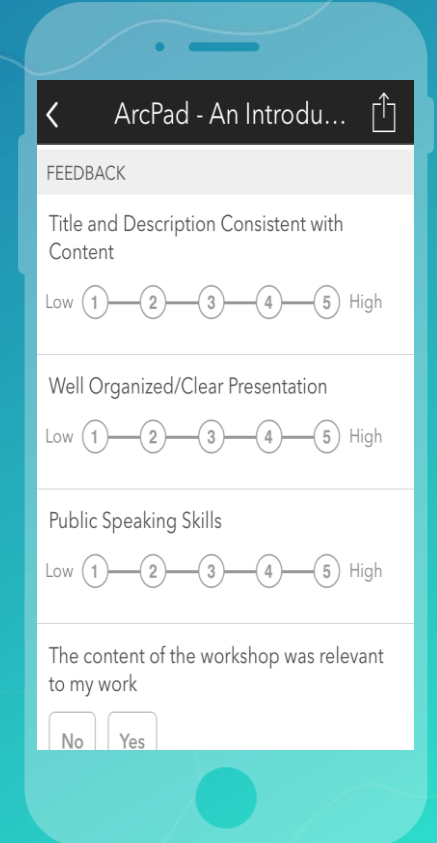
Select the session you attended



Scroll down to find the survey



Complete Answers and Select "Submit"



Questions?





esri

THE
SCIENCE
OF
WHERE