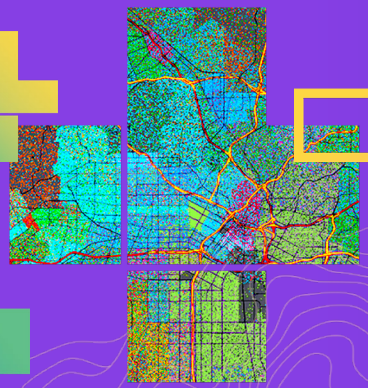
Big Data and ArcGIS: An Introduction to ArcGIS GeoAnalytics Server

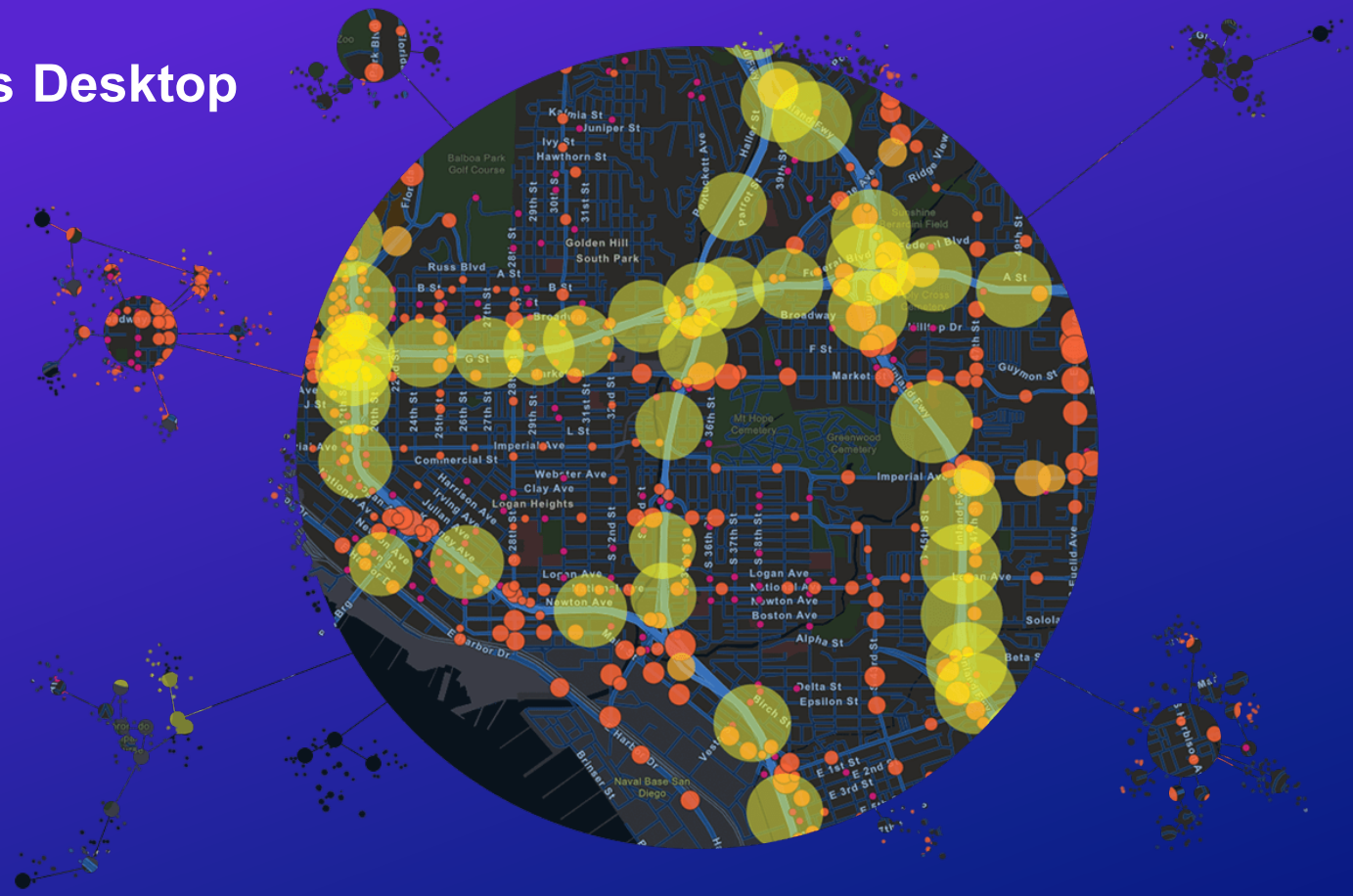
Sarah Ambrose + Noah Slocum

2020 ESRI DEVELOPER SUMMIT | Palm Springs, CA



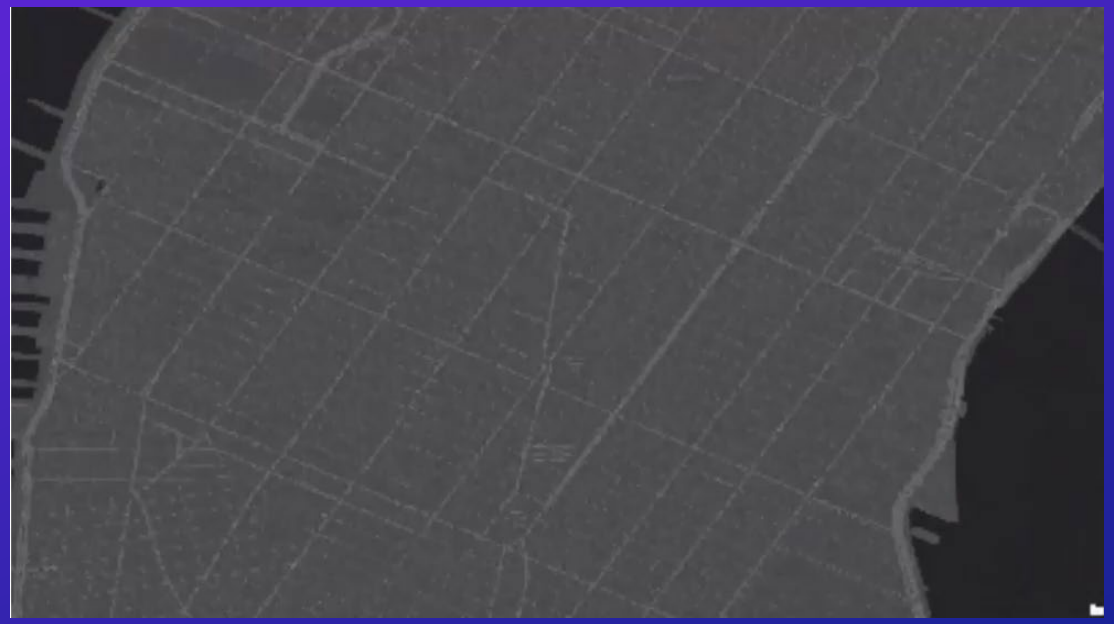
Agenda

- What is “GeoAnalytics”?
- GeoAnalytics Server vs. GeoAnalytics Desktop
- Analysis Tools
- Data Integration
- System Requirements
- Questions



Every solution starts with a problem….

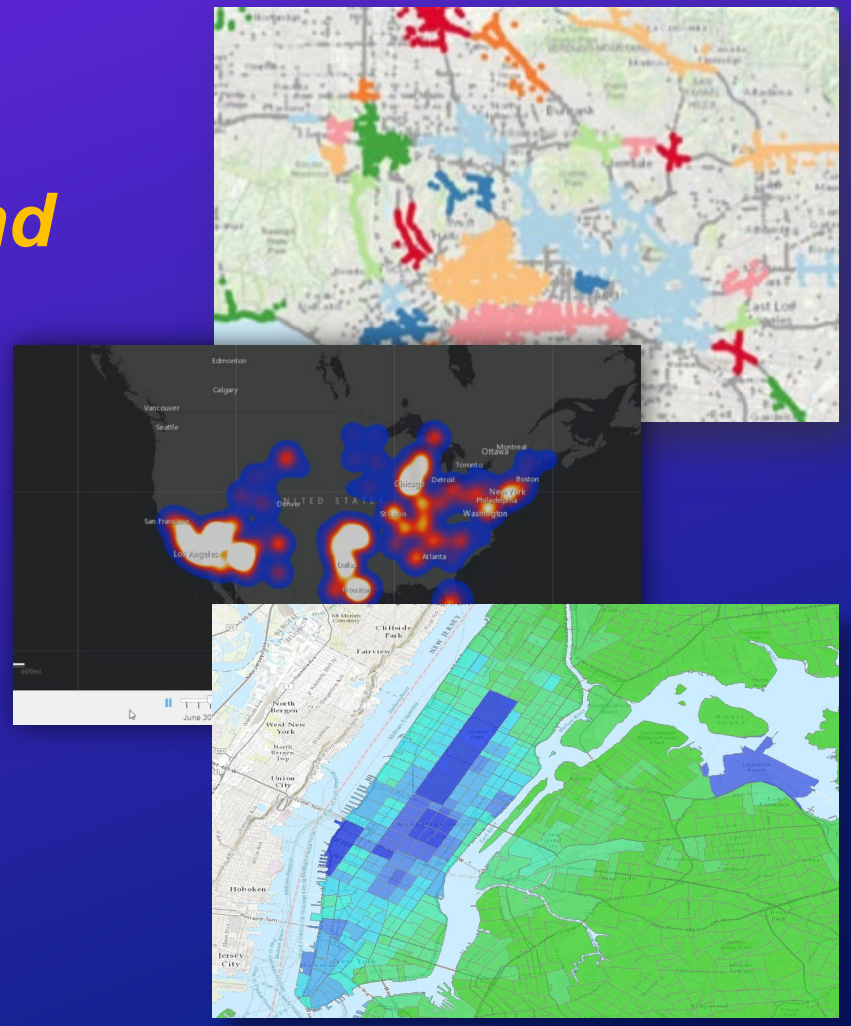
VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	store_and_fwd_flag	RatecodeID	PULocationID	DOLocationID	passenger_count	trip_distance	fare_amount	extra	mta_tax	tip_amount	tolls_amount	ehail_fee	improvement_surcharge	total_amount	payment_type	trip_type
2	1/1/2018 0:18	1/1/2018 0:24	N	1	236	236	5	0.7	6	0.5	0.5	0	0		0.3	7.3	2	1
2	1/1/2018 0:30	1/1/2018 0:46	N	1	43	42	5	3.5	14.5	0.5	0.5	0	0		0.3	15.8	2	1
2	1/1/2018 0:07	1/1/2018 0:19	N	1	74	152	1	2.14	10	0.5	0.5	0	0		0.3	11.3	2	1
2	1/1/2018 0:32	1/1/2018 0:33	N	1	255	255	1	0.03	-3	-0.5	-0.5	0	0		-0.3	-4.3	3	1
2	1/1/2018 0:32	1/1/2018 0:33	N	1	255	255	1	0.03	3	0.5	0.5	0	0		0.3	4.3	2	1
2	1/1/2018 0:38	1/1/2018 1:08	N	1	255	161	1	5.63	21	0.5	0.5	0	0		0.3	22.3	2	1
2	1/1/2018 0:18	1/1/2018 0:28	N	1	189	65	5	1.71	8.5	0.5	0.5	0	0		0.3	9.8	2	1
2	1/1/2018 0:38	1/1/2018 0:55	N	1	189	225	5	3.45	14.5	0.5	0.5	3.16	0		0.3	18.96	1	1
2	1/1/2018 0:05	1/1/2018 0:18	N	1	129	82	1	1.61	10	0.5	0.5	0	0		0.3	11.3	2	1
2	1/1/2018 0:35	1/1/2018 0:42	N	1	226	7	1	1.87	7.5	0.5	0.5	0	0		0.3	8.8	2	1
2	1/1/2018 0:21	1/1/2018 0:39	N	1	145	129	2	4.12	16.5	0.5	0.5	0	0		0.3	17.8	2	1
2	1/1/2018 0:56	1/1/2018 1:04	N	1	7	223	2	1.22	7	0.5	0.5	0	0		0.3	8.3	2	1
2	1/1/2018 0:11	1/1/2018 0:30	N	1	255	189	1	4.67	17	0.5	0.5	0	0		0.3	18.3	2	1
2	1/1/2018 0:57	1/1/2018 1:12	N	1	97	188	1	2.71	11.5	0.5	0.5	3.84	0		0.3	16.64	1	1
2	1/1/2018 0:36	1/1/2018 0:51	N	1	244	75	2	6.01	19	0.5	0.5	4	0		0.3	24.3	1	1
1	1/1/2018 0:07	1/1/2018 0:15	N	1	225	37	1	1.9	8	0.5	0.5	3	0		0.3	12.3	1	1
1	1/1/2018 0:25	1/1/2018 0:42	N	1	36	145	2	4.3	15.5	0.5	0.5	3.35	0		0.3	20.15	1	1
1	1/1/2018 0:42	1/1/2018 1:00	N	1	145	173	1	6.9	22	0.5	0.5	0	0		0.3	23.3	1	1
2	1/1/2018 0:06	1/1/2018 0:08	N	1	49	49	1	0.3	3.5	0.5	0.5	0	0		0.3	4.8	2	1
2	1/1/2018 0:34	1/1/2018 0:52	N	1	40	113	1	4.47	16.5	0.5	0.5	3.56	0		0.3	23.31	1	1
1	1/1/2018 0:25	1/1/2018 0:28	N	1	179	7	1	0.5	4.5	0.5	0.5	0	0		0.3	5.8	1	1
2	1/1/2018 0:36	1/1/2018 0:51	N	1	7	193	1	1.82	9	0.5	0.5	0	0		0.3	10.3	1	1
2	1/1/2018 0:53	1/1/2018 1:26	N	1	97	74	1	11.79	36	0.5	0.5	7.46	0		0.3	46.71	1	1
1	1/1/2018 0:11	1/1/2018 0:22	N	1	255	112	1	1.9	9	0.5	0.5	3.05	0		0.3	13.35	1	1
1	1/1/2018 0:40	1/1/2018 1:01	N	1	255	28	1	10.3	29	0.5	0.5	5	0		0.3	35.3	1	1
2	1/1/2018 0:15	1/1/2018 0:25	N	1	80	80	1	1.66	8.5	0.5	0.5	1.96	0		0.3	11.76	1	1
2	1/1/2018 0:35	1/1/2018 0:48	N	1	255	232	1	2.91	12	0.5	0.5	3.32	0		0.3	16.62	1	1
2	1/1/2018 0:55	1/1/2018 1:28	N	1	256	50	1	6.09	25	0.5	0.5	6.58	0		0.3	32.88	1	1
2	1/1/2018 0:41	1/1/2018 0:56	N	1	179	75	5	5.3	17	0.5	0.5	3.66	0		0.3	21.96	1	1
2	1/1/2018 0:36	1/1/2018 0:44	N	1	41	75	1	1.63	8	0.5	0.5	1.86	0		0.3	11.16	1	1
2	1/1/2018 0:48	1/1/2018 0:51	N	1	75	74	1	0.91	4.5	0.5	0.5	0	0		0.3	5.8	2	1
2	1/1/2018 0:56	1/1/2018 1:00	N	1	74	74	2	0.92	5	0.5	0.5	0	0		0.3	6.3	1	1
2	1/1/2018 0:27	1/1/2018 0:34	N	1	7	223	1	0.98	6	0.5	0.5	0	0		0.3	7.3	2	1
2	1/1/2018 0:41	1/1/2018 0:52	N	1	179	7	1	1.42	9	0.5	0.5	0	0		0.3	10.3	2	1
2	1/1/2018 0:48	1/1/2018 0:53	N	1	254	254	1	0.42	5	0.5	0.5	0	0		0.3	6.3	1	1
2	1/1/2018 0:46	1/1/2018 1:02	N	1	255	36	1	3.15	13	0.5	0.5	2.86	0		0.3	17.16	1	1
2	1/1/2018 1:03	1/1/2018 1:09	N	1	260	226	1	0.85	5.5	0.5	0.5	0	0		0.3	6.8	2	1
2	1/1/2018 0:30	1/1/2018 0:38	N	1	181	190	5	1.28	7.5	0.5	0.5	1.76	0		0.3	10.56	1	1
2	1/1/2018 0:39	1/1/2018 0:48	N	1	190	89	5	1.58	8	0.5	0.5	1.86	0		0.3	11.16	1	1
2	1/1/2018 0:59	1/1/2018 1:25	N	1	188	76	5	4.43	18.5	0.5	0.5	0	0		0.3	19.8	1	1



How do I make sense of large amounts of data?

What is GeoAnalytics?

GeoAnalytics *parallelizes computing* to quickly analyze large amounts of *vector and tabular data*



A collection of analysis tools to identify *patterns*, *relationships*, *anomalies* and *incidents* in large amounts of data across space and time

Why use GeoAnalytics?

- My existing tools and workflows aren’t processing my data fast enough
- My data is growing and I need a better way of managing and analyzing it
- I need to distill my data into something more manageable to use in other analysis
- My data has a ton of noise and I want to explore it to bring out what’s really important

Powerful Analytics

Leverage spatiotemporal analysis to answer questions like:

- Using billions of emergency calls accumulated over the past decade, which counties have the largest number of calls per square mile?
- Which stationary pressure sensors in my pipe network have experienced anomalous events in the past 24 hours? Where are there hot spots of anomalous events?
- Where have my delivery trucks traveled and where is the highest density of unique delivery truck paths? Where do delivery trucks travel the slowest?

GeoAnalytics is available through...

GeoAnalytics Server

Distributed processing across multiple *server cores and machines* with ArcGIS Enterprise

Requirements: Enterprise + GeoAnalytics Server License

GeoAnalytics Desktop

Parallel processing across cores on your laptop or desktop with ArcGIS Pro

Requirements: Advanced License

When to use Desktop or Server for GeoAnalytics

- Use **GeoAnalytics Server** when you want to:
  - Bring big data analysis to your entire organization
  - Leverage the power of one or multiple server machines
  - Connect to external big data storage and existing web layers
  - Extend using custom analysis.

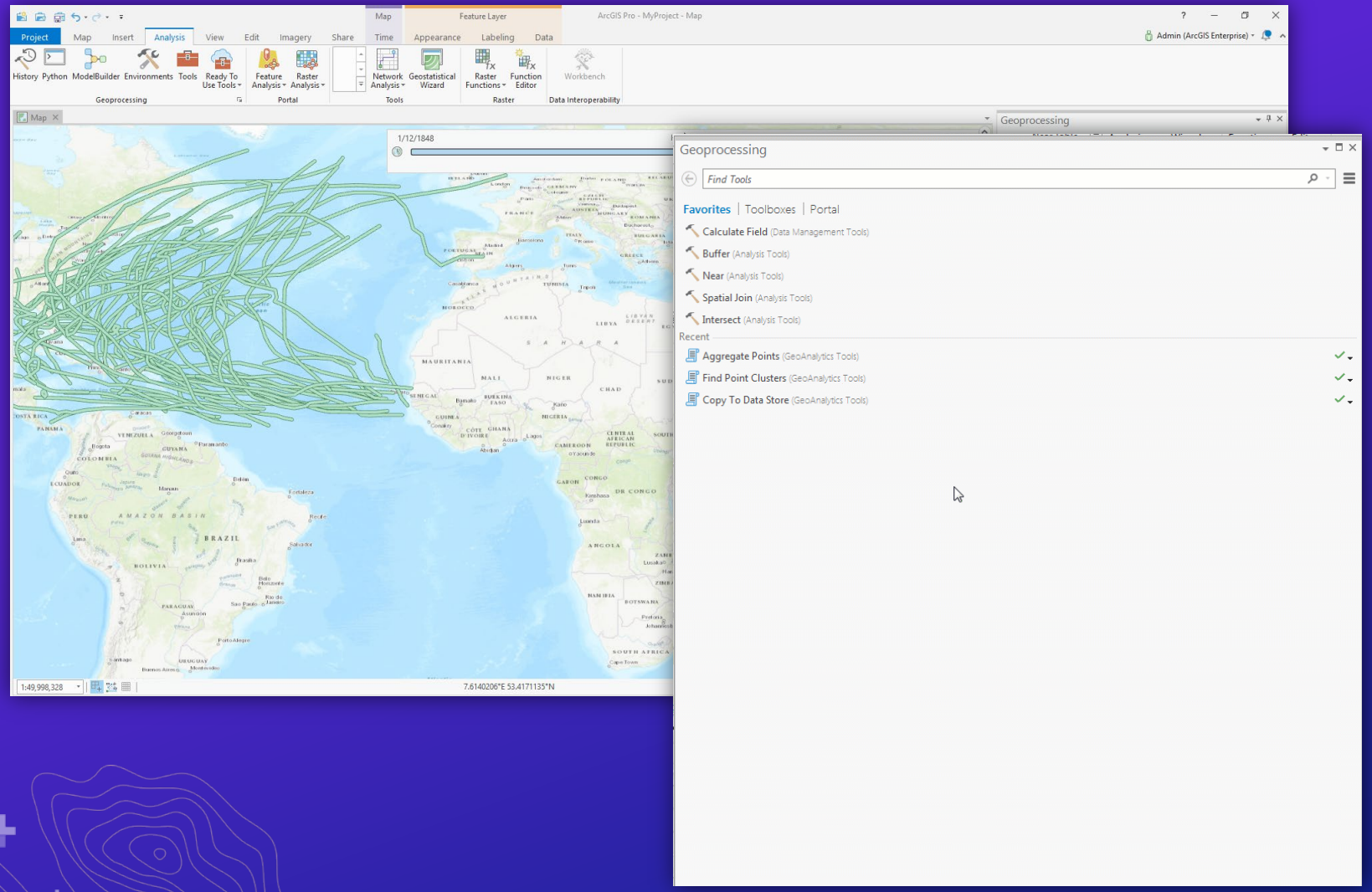
- Use **GeoAnalytics Desktop** when you want to:
  - Process local data (from files, databases) faster than before on your own desktop machine
  - Prototyping workflows you want to use with GeoAnalytics Server.

When to use Desktop or Server for GeoAnalytics

	GeoAnalytics Server	GeoAnalytics Desktop
Input data	- Big data file shares * - Hosted feature layers - Feature services	- File geodatabase - Enterprise geodatabase - Shapefiles
Output data	- Hosted feature layers - Big data file shares *	- File geodatabase - Enterprise geodatabase - Shapefiles
Scaling out analysis	- Control the number of machines - Control the percentage of cores and RAM - Scale out data storage with spatiotemporal data store	- One machine only - Control the percentage of RAM (or a value)
Tools	- 26 tools (and adding more) - Run Python Script	- 19 tools (and adding more)
Interface	- REST + the ArcGIS API for Python - Pro and Arcpy (and model builder) - Portal Map Viewer	- Pro and Arcpy (and model builder)

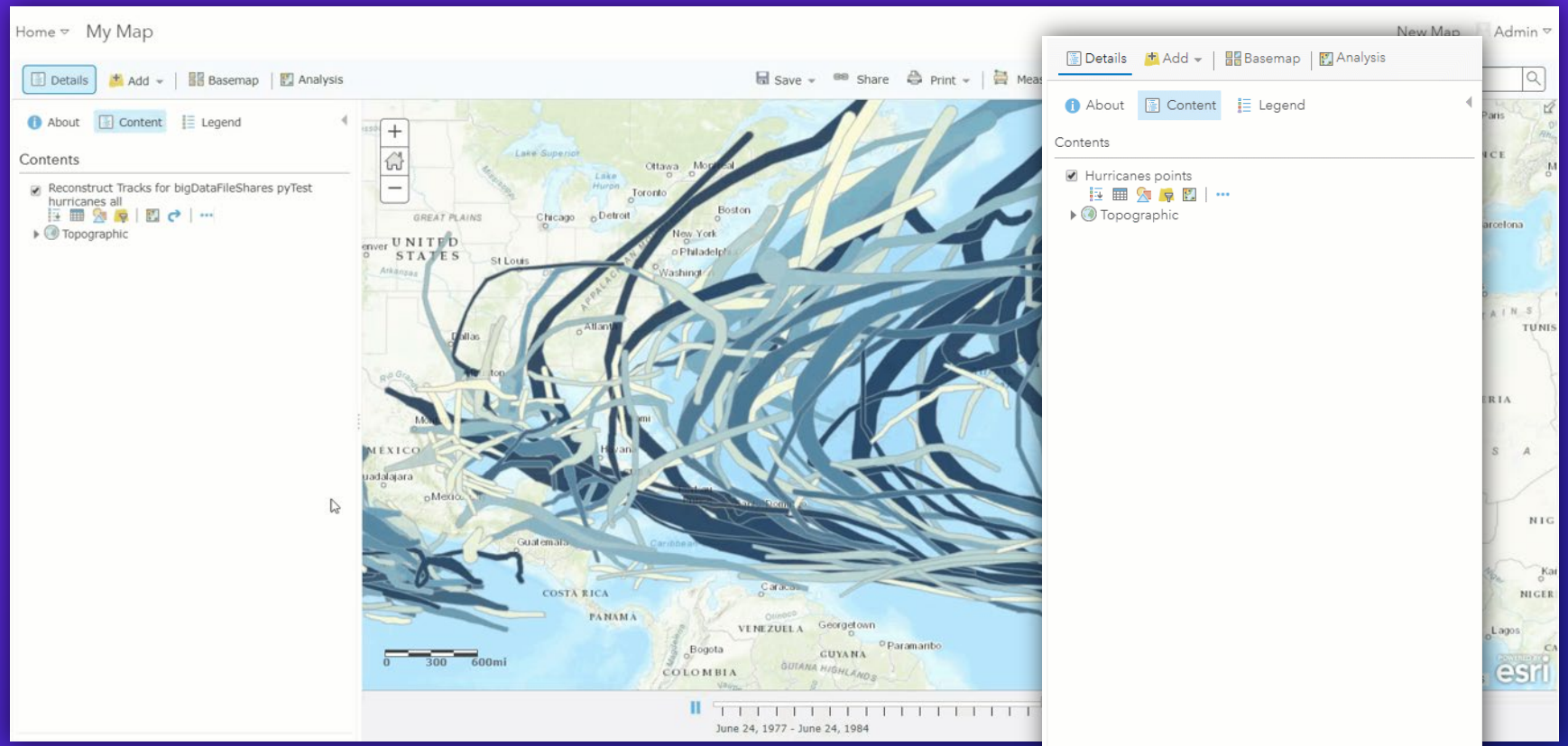
Blog post covering this topic

GeoAnalytics Server is ready to use with:



ArcGIS Pro

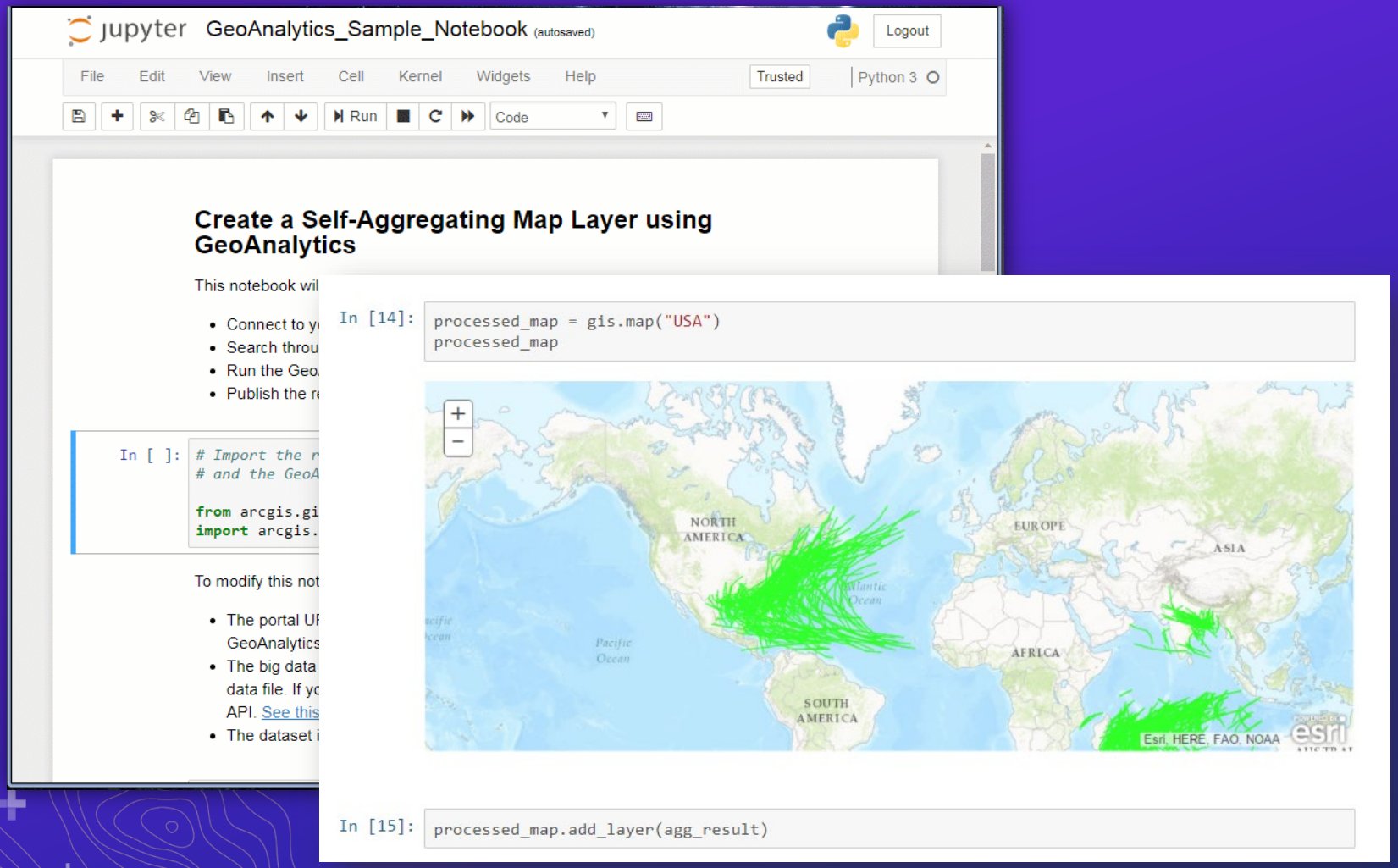
GeoAnalytics Server is ready to use with:



ArcGIS Pro

Enterprise portal

GeoAnalytics Server is ready to use with:

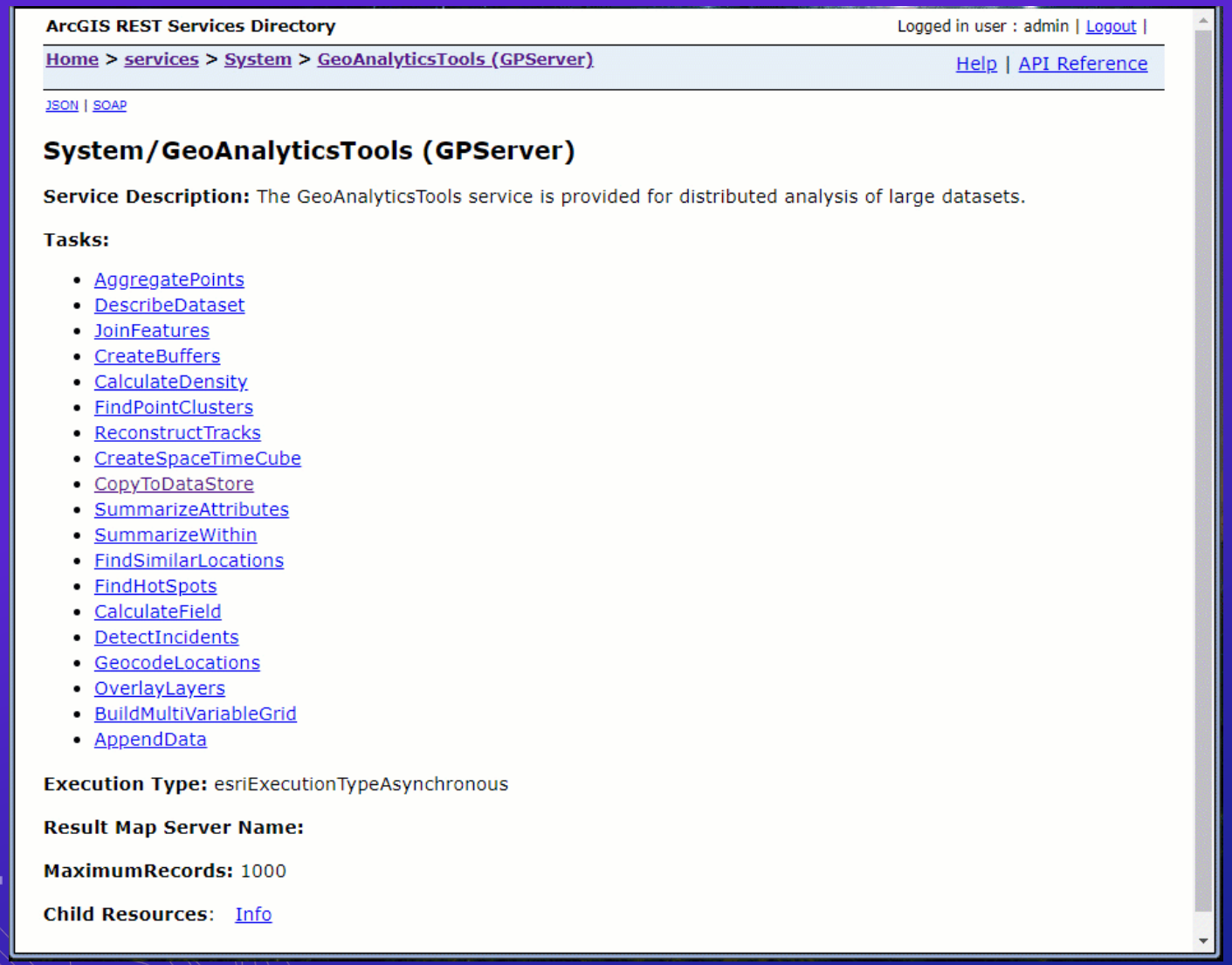


ArcGIS Pro

Enterprise portal

ArcGIS API for Python

GeoAnalytics Server is ready to use with:



ArcGIS REST Services Directory

Logged in user : admin | Logout |

Home > services > System > GeoAnalyticsTools (GPServer)

Help | API Reference

JSON | SOAP

System/GeoAnalyticsTools (GPServer)

Service Description: The GeoAnalyticsTools service is provided for distributed analysis of large datasets.

Tasks:

- AggregatePoints
- DescribeDataset
- JoinFeatures
- CreateBuffers
- CalculateDensity
- FindPointClusters
- ReconstructTracks
- CreateSpaceTimeCube
- CopyToDataStore
- SummarizeAttributes
- SummarizeWithin
- FindSimilarLocations
- FindHotSpots
- CalculateField
- DetectIncidents
- GeocodeLocations
- OverlayLayers
- BuildMultiVariableGrid
- AppendData

Execution Type: esriExecutionTypeAsynchronous

Result Map Server Name:

MaximumRecords: 1000

Child Resources: Info

ArcGIS Pro

Enterprise portal

ArcGIS API for Python

ArcGIS REST API

Demo # 1

Noah Slocum

Analysis Capabilities

Analysis Capabilities

Summarize Data

- Aggregate Points
- Build Multi-Variable Grid
- Describe Dataset
- Join Features
- Reconstruct Tracks
- Summarize Attributes
- Summarize Within

Analyze Patterns

- Calculate Density
- Create Space Time Cube
- Find Hot Spots
- Find Point Clusters *+
- Forest-based Classification and Regression
- Generalized Linear Regression

Find Locations

- Detect Incidents
- Find Similar Locations
- Find Dwell Locations +
- Geocode Locations

Use Proximity

- Create Buffers

Data Enrichment

- Enrich from Multi-Variable Grid

Manage Data

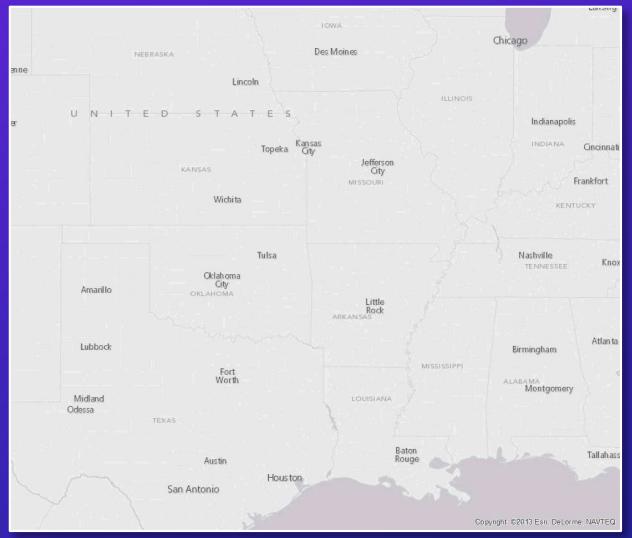
- Append Data
- Calculate Field
- Clip Layer
- Copy to Data Store
- Dissolve Boundaries
- Merge Layers
- Overlay Layers

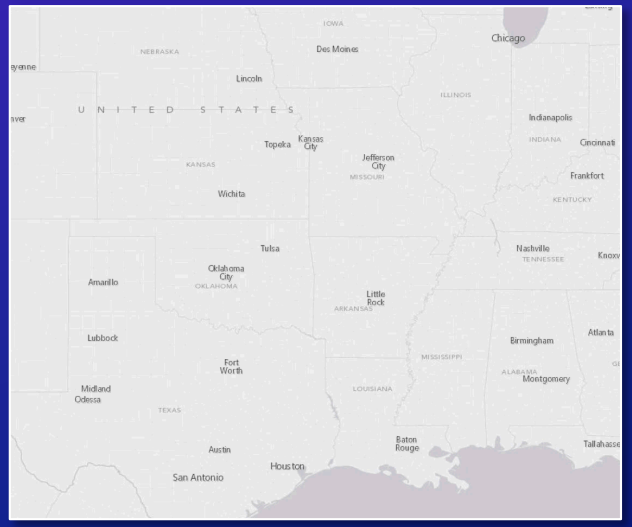
+ New at 10.8

Analysis Capabilities + Space and Time

Work with data in both space and time

- Use GeoAnalytics to perform spatiotemporal analysis
- Define your temporal input data:
  - Instants (a moment in time)
  - Intervals (a duration in time)
- Analyze data in time steps
- Visualize results across time using Pro and Map Viewer



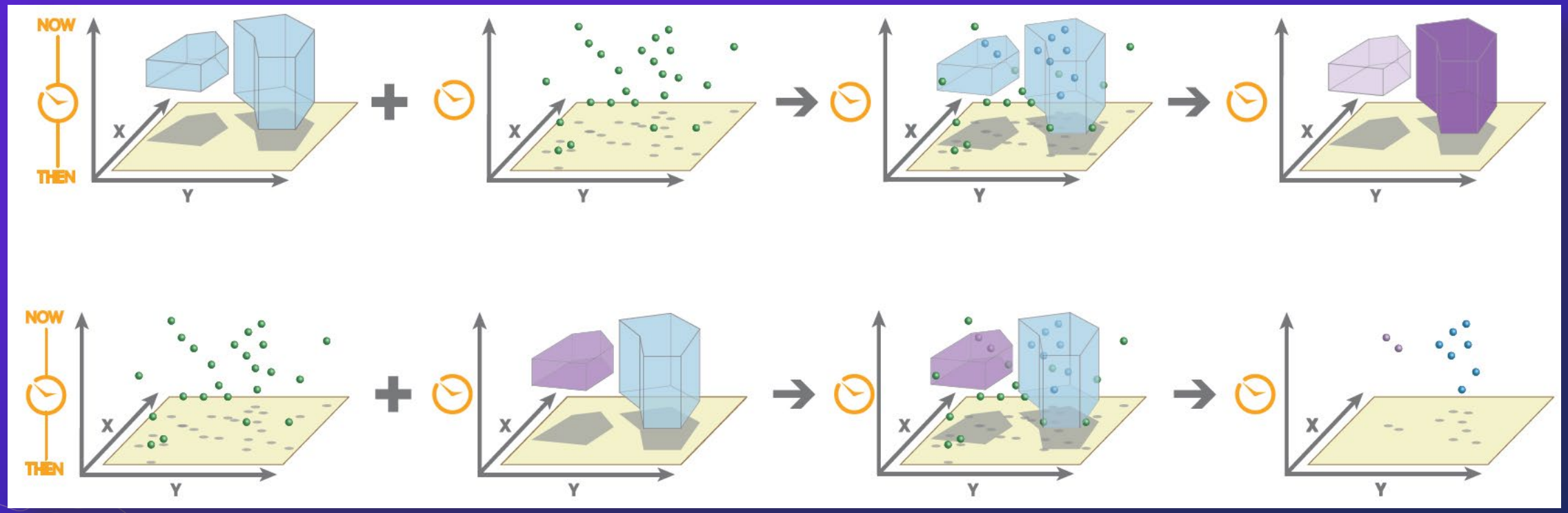


Join Features

Join data by location, time, and attributes

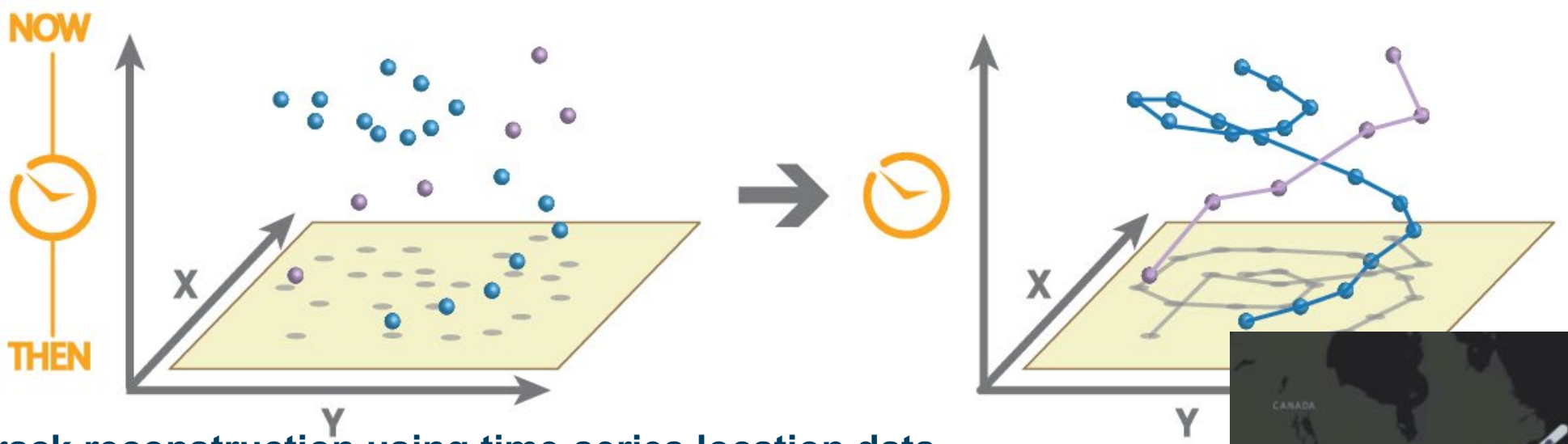
(a)

(b)



Space-time join of (a) polygon to point features and (b) point to polygon features.

Reconstruct Tracks



Track reconstruction using time-series location data

- Track data examples:
  - Hurricanes
  - Ships
  - Airplanes

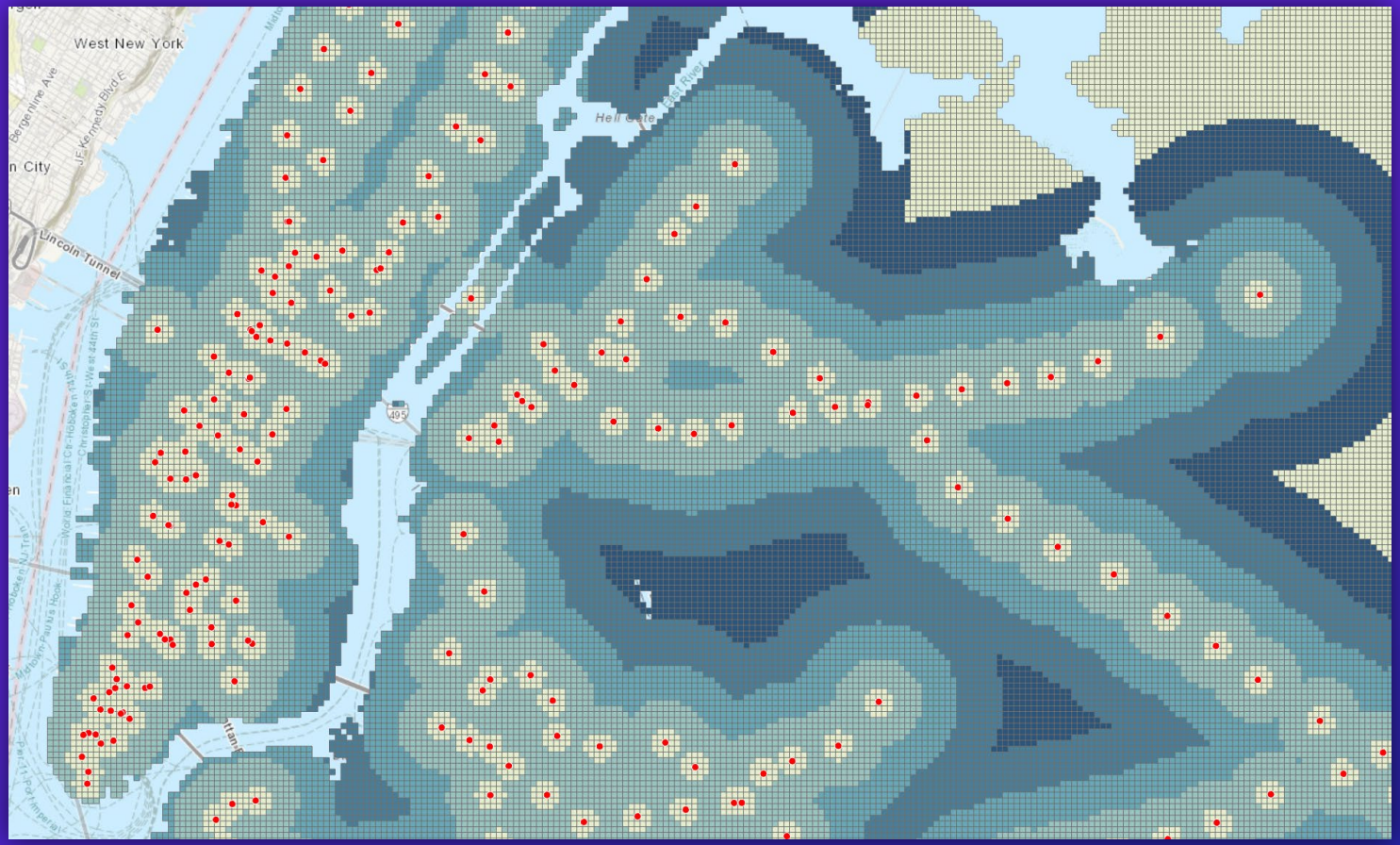
Build Multi-Variable Grid

- Built to aggregate multiple datasets into one by calculating:
  - Distance to nearest
  - Attributes of nearest
  - Summary of intersecting
  - Summary within a given distance

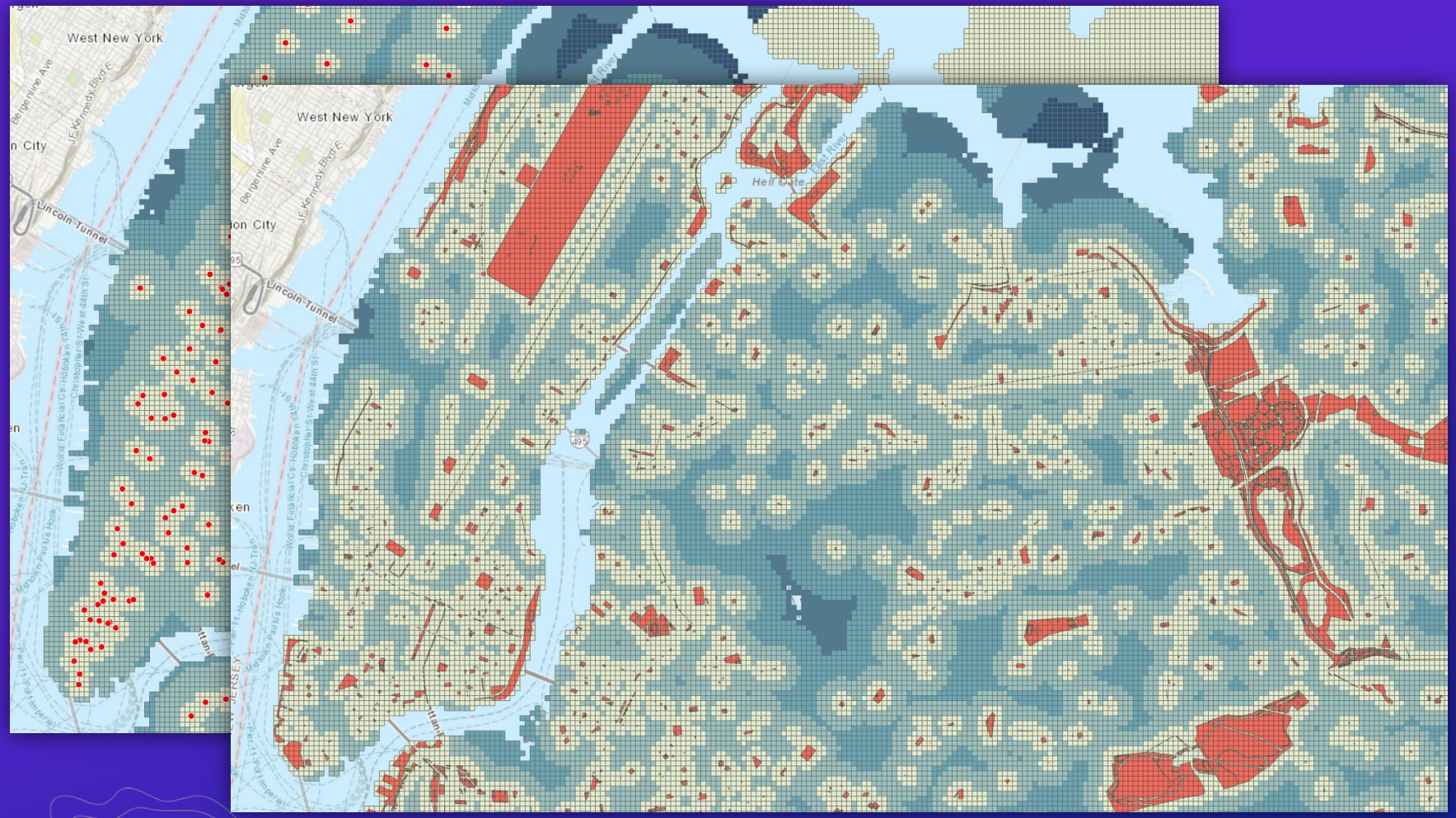
  …on one or more layers of interest

- Use this layer as:
  - Visualization
  - Input to statistical tools (like GLR + Forest-based Classification and Regression)
  - An enrichment layer

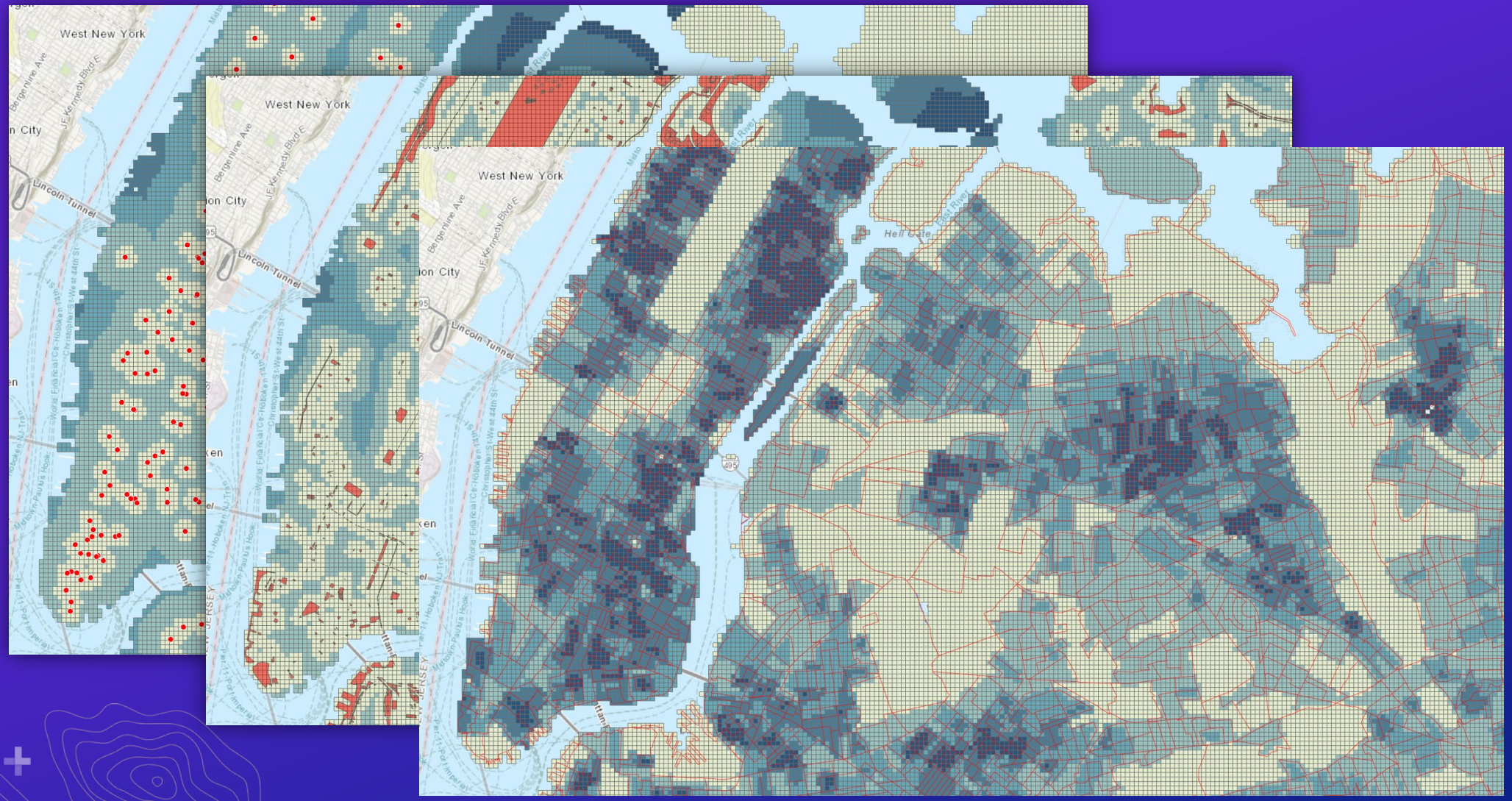
Build Multi-Variable Grid – Nearest Subway Station



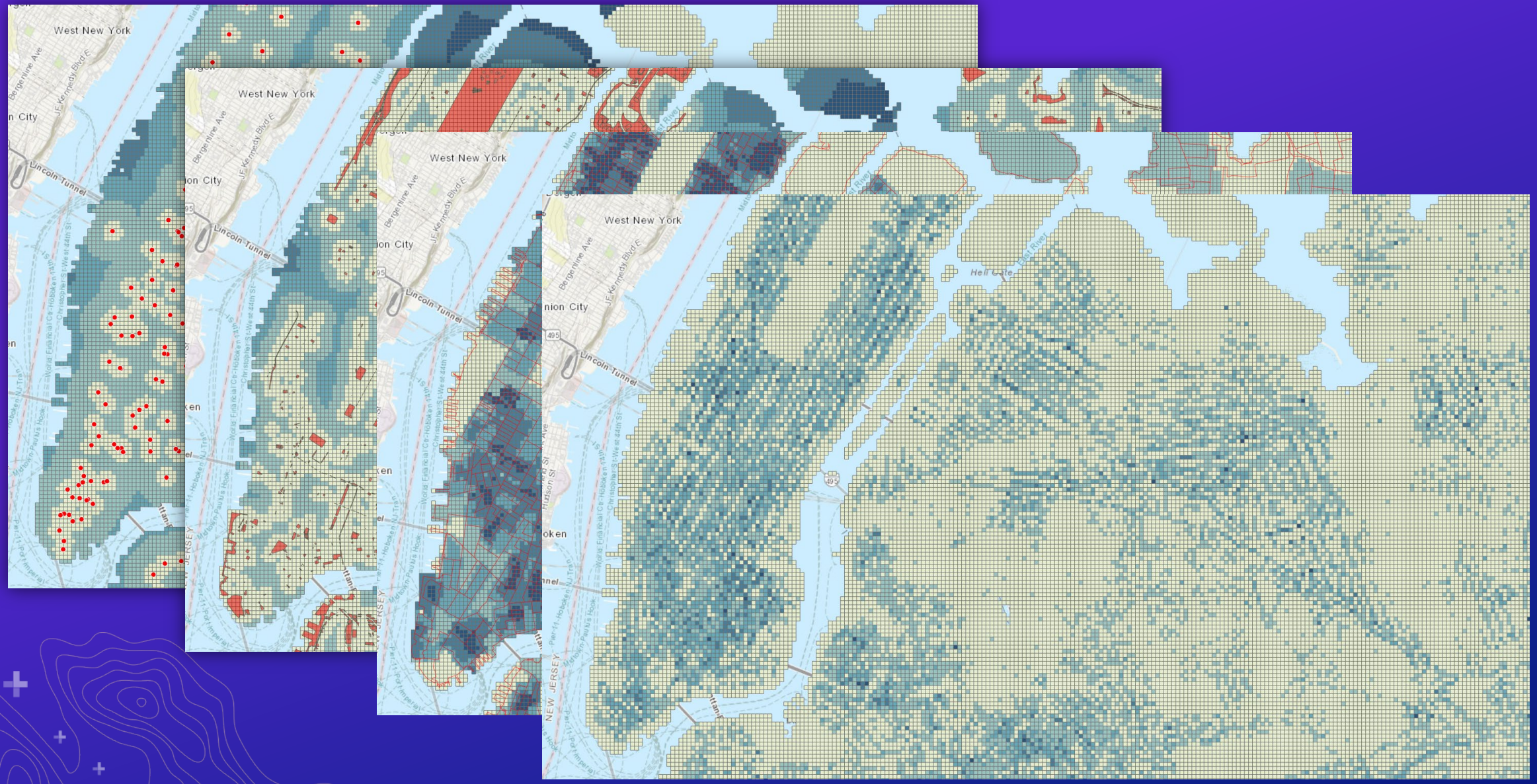
Build Multi-Variable Grid – Nearest Park



Build Multi-Variable Grid – Population per Sq. Mile



Build Multi-Variable Grid – Number of Complaints



Access and use PySpark with GeoAnalytics Server

Use Run Python Script to execute distributed analysis

- Run a custom python script on your GeoAnalytics Server site
- Use other python functionality and distribute analysis across your site
- Create an analysis pipeline to chain GeoAnalytics tools together
- Use pyspark (ml, sql) and data frames

Demo # 2

Noah Slocum

Data Integration

GeoAnalytics Server

What types of data can I analyze? And from where?

- Access and share data within Enterprise with *your Enterprise portal*
- Seamlessly analyze data collected with *ArcGIS GeoEvent Server*
- Analyze data in *Hive*, *HDFS*, and *files*
- Connect to Azure and Amazon *cloud stores*

Easily Connect To Your Big Data

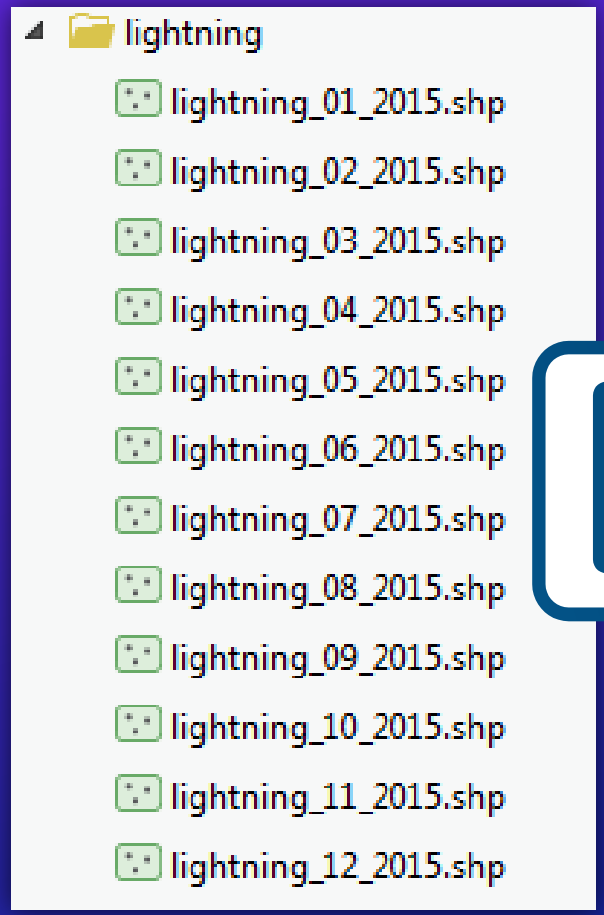
Big Data File Shares

Read directly from files stored in

- Hive
- HDFS
- Shared folders
- Cloud stores

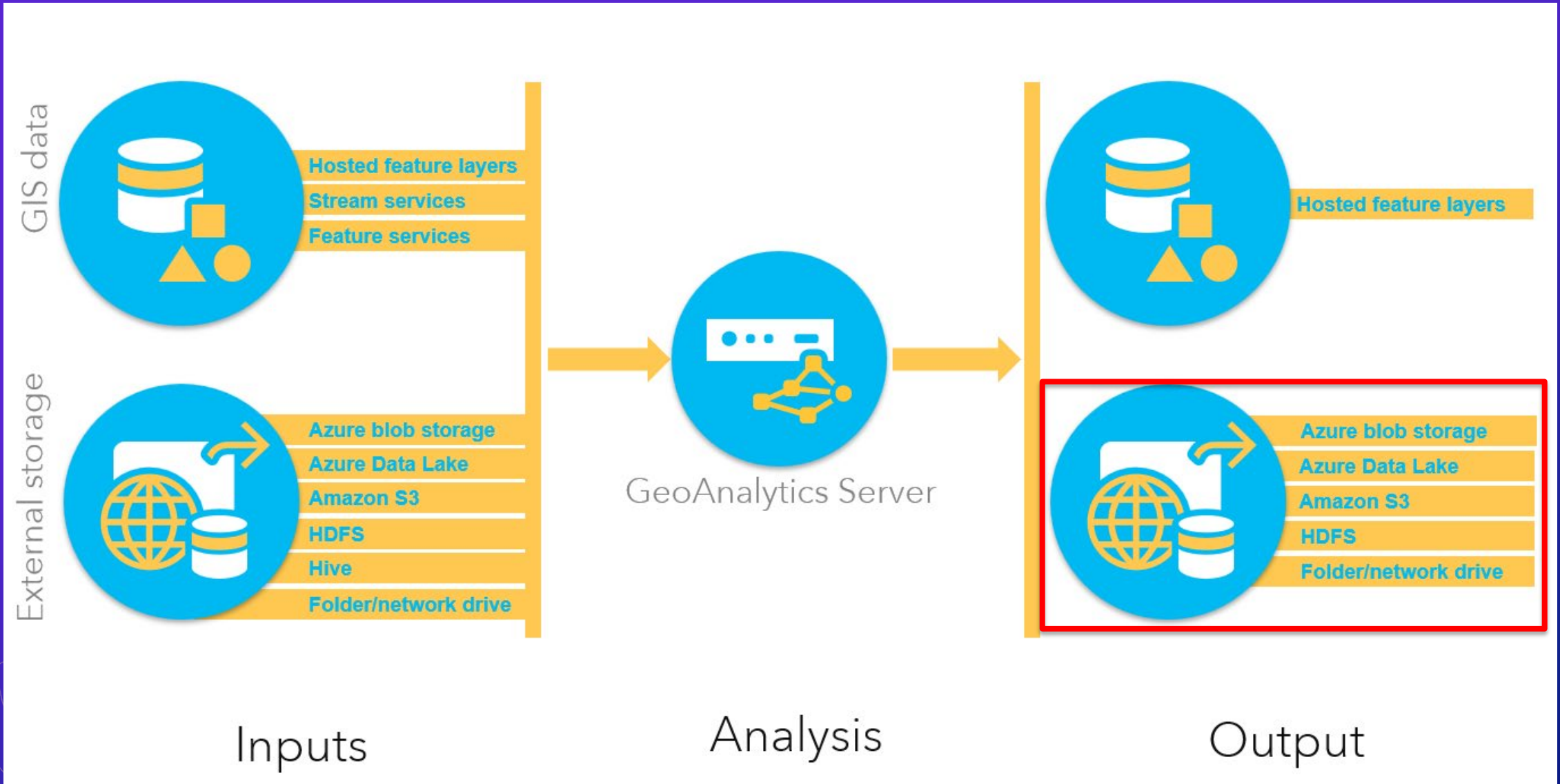
Supported file types include

- Delimited files
- Shapefiles
- ORC
- Parquet

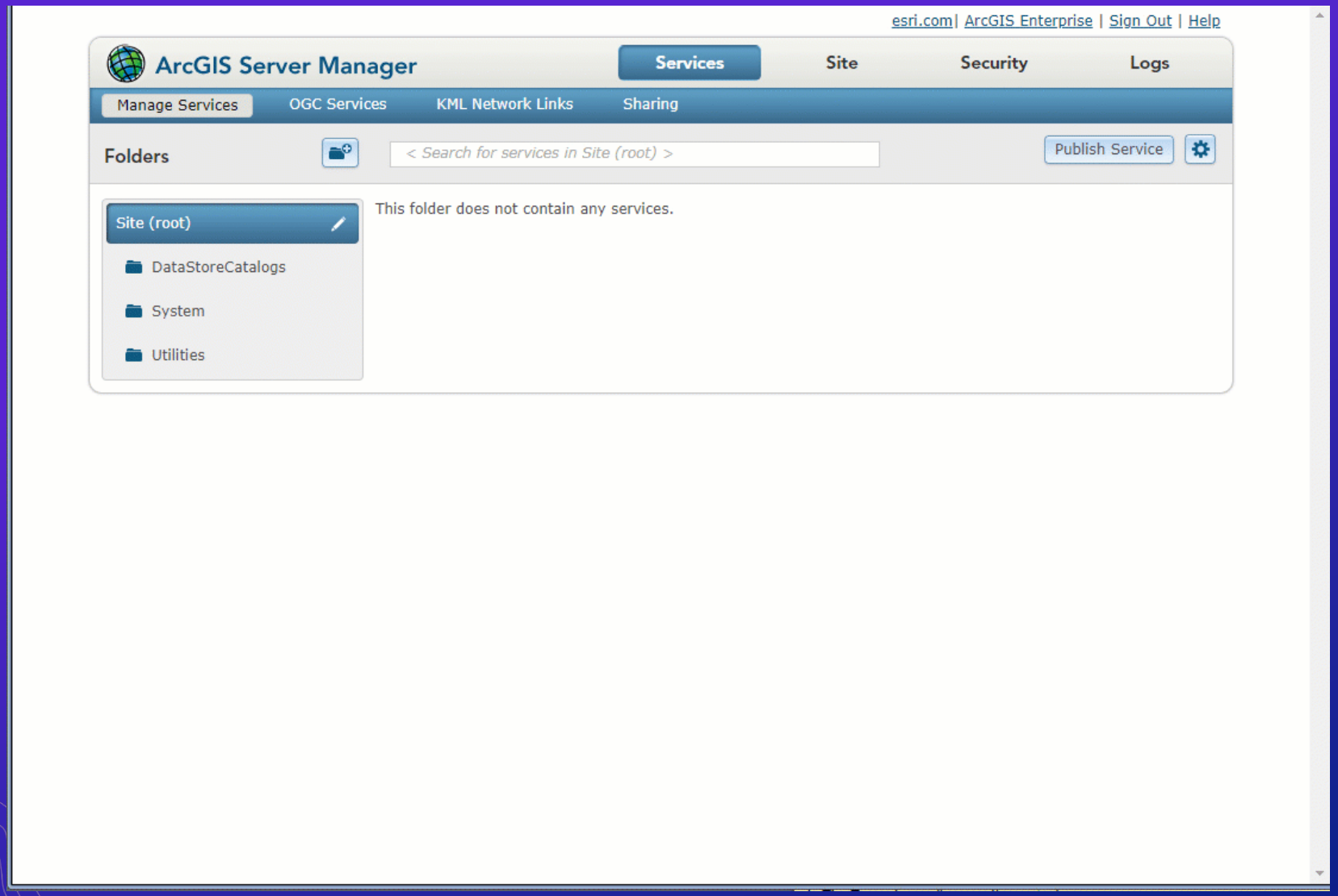


Use multiple files with a matching schema as a single dataset

Write to big data file shares



How do I actually use all those data sources?

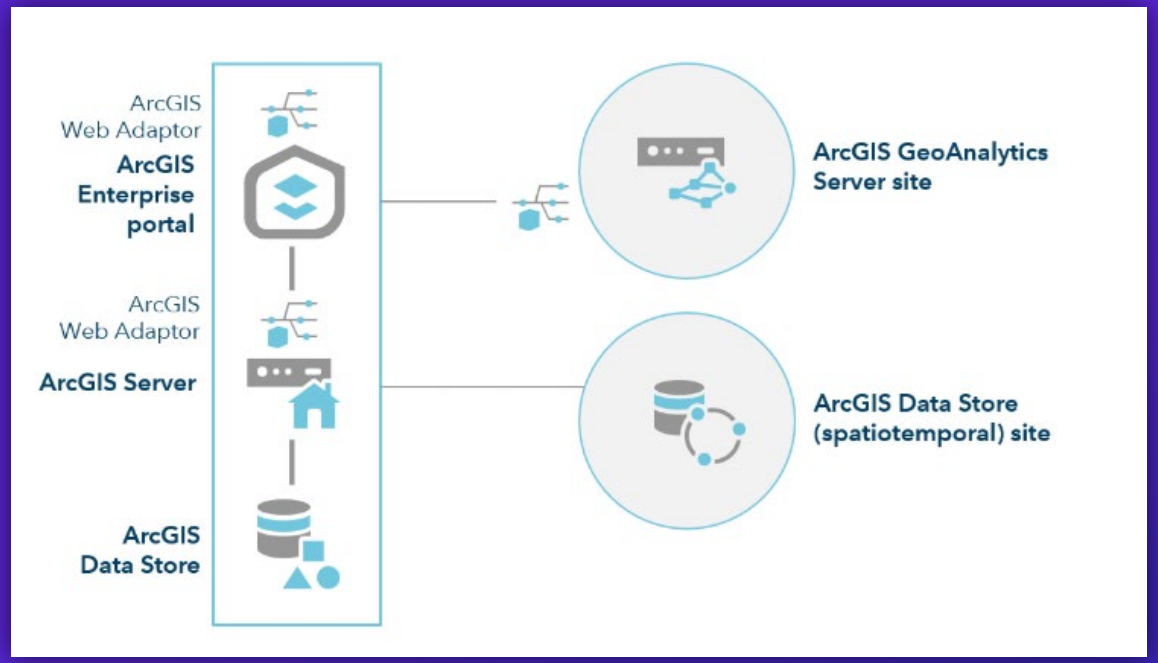


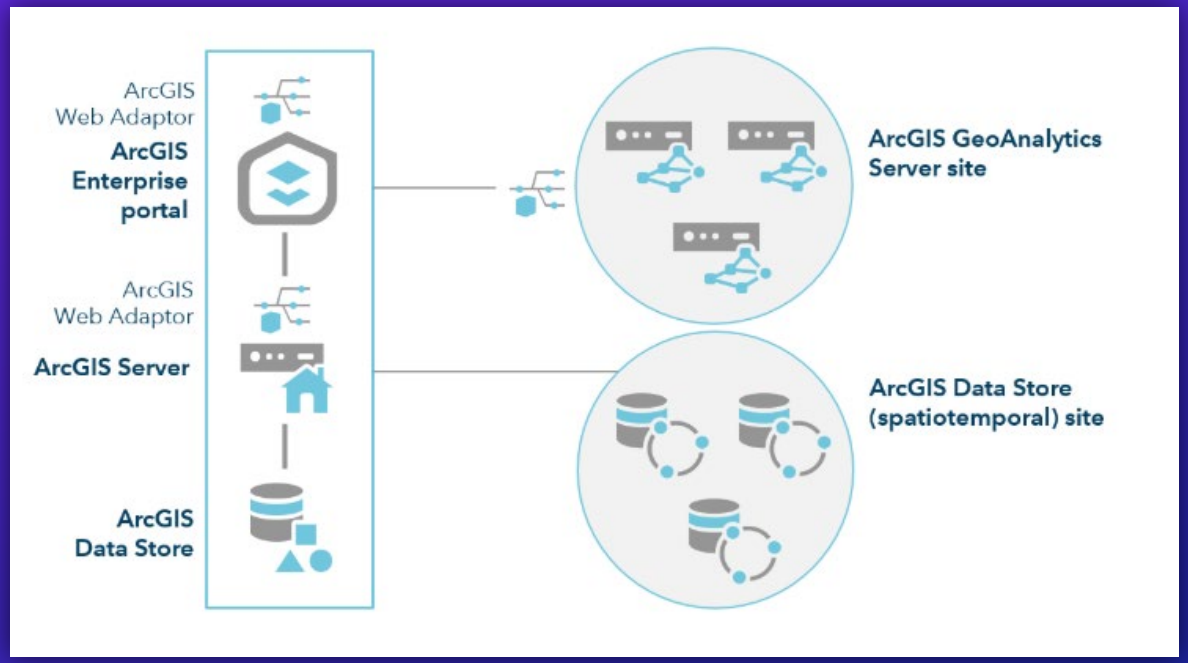
Deploying GeoAnalytics Server

What do I need to install to use ArcGIS GeoAnalytics Server?

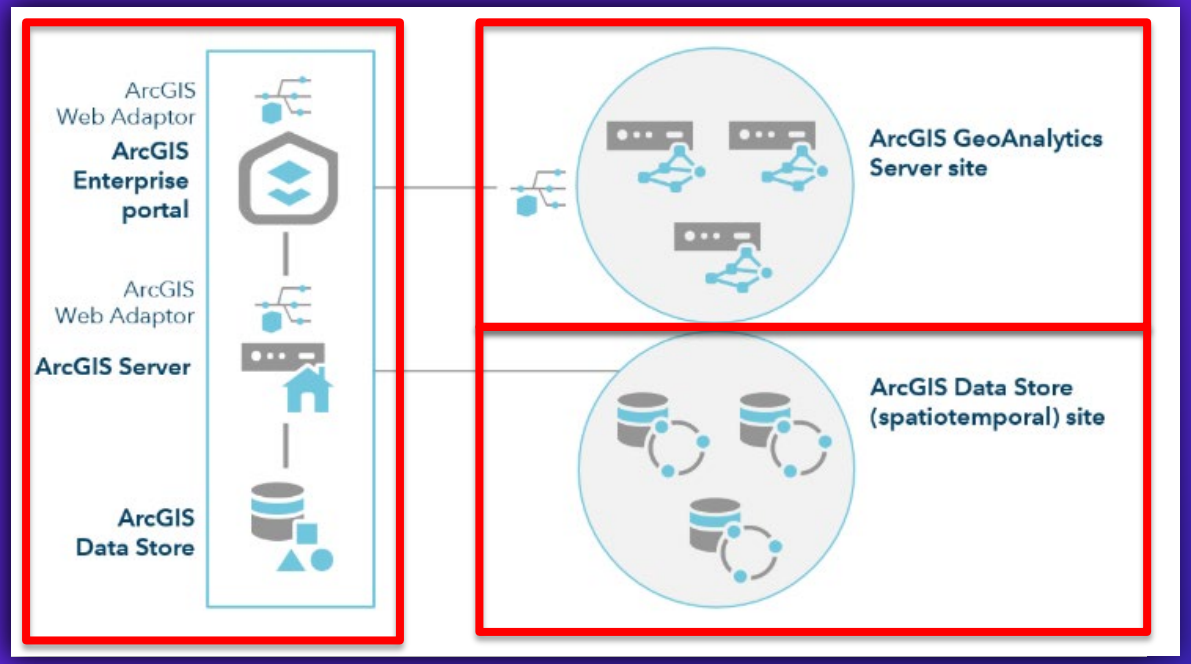
- Portal for ArcGIS
- ArcGIS Web Adaptor
- ArcGIS Server
- ArcGIS Data Store (Relational + Spatiotemporal)

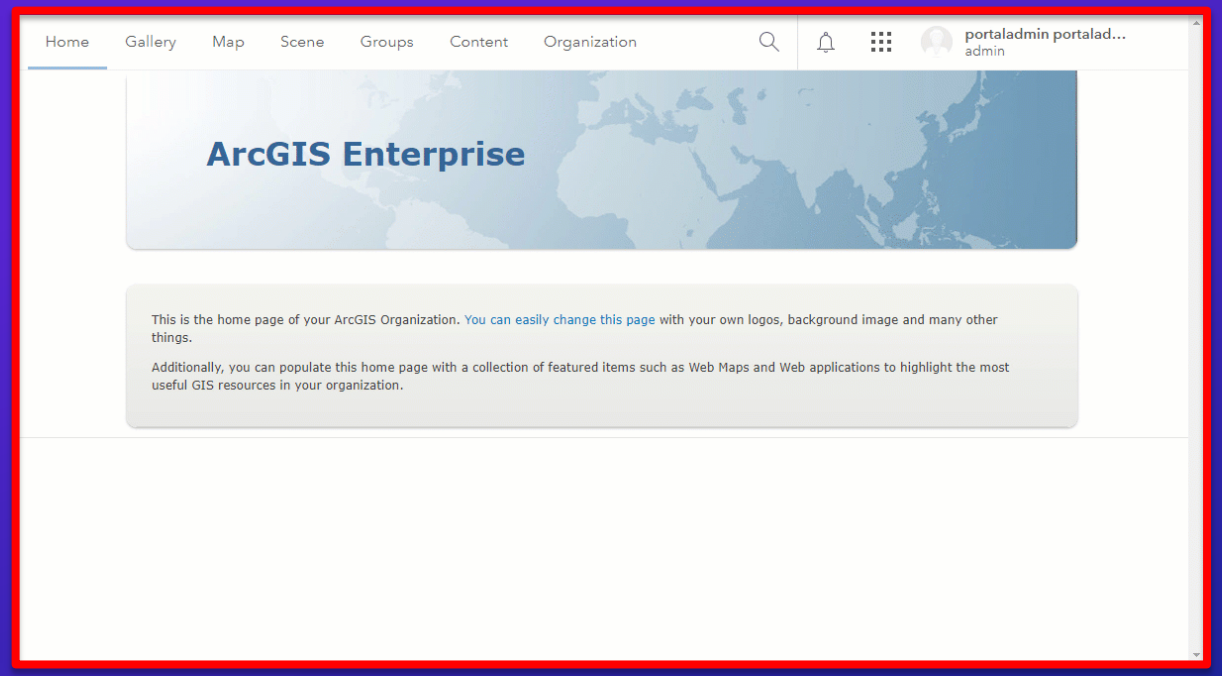
GeoAnalytics Sites | *1 or 3 machines*





GeoAnalytics Sites | 1 or 3 machines





Steps:

1. Set up your Base Deployment
2. Set up GeoAnalytics Server, license it as GeoAnalytics, set up a web adaptor, federate with your portal
3. Install ArcGIS Data Store and configure the spatiotemporal data store with you hosting server
4. Set your GeoAnalytics Server as your GeoAnalytics Server

System Requirements

- **Base ArcGIS Enterprise deployment:**
  - Minimum 4 cores and 16 GB RAM

- **ArcGIS Spatiotemporal Big Data Store:**
  - Minimum 16 GB RAM for each machine (**32GB Recommended**)
  - Fast disk/sufficient disk space for data
  - At least the same number of machines as your GeoAnalytics Server

- **ArcGIS GeoAnalytics Server (1 or 3 machines):**
  - Minimum 4 cores and 16 GB RAM for each machine
  - Sufficient temp space for compute

Deployment

- **Compatible with both Windows and Linux systems**
- **Can be deployed on premises or in the cloud**

- **Simplify deployment of ArcGIS Enterprise and GeoAnalytics Server with:**
  - **ArcGIS Enterprise Builder –** single-machine base deployment
  - **Chef Cookbooks for ArcGIS –** multimachine and HA deployments
  - **ArcGIS Enterprise Cloud Builder –** AWS and Azure

Summary

- *Integrated*: Works with your existing big data storage AND/OR existing GIS data AND/OR what you currently use (Desktop or Enterprise).
- *Spatiotemporal*: Tools are designed to analyze data in space and time.
- *Accelerated*: Speeds up analytical processing time using built-in parallel compute.
- *Actionable*: Able to crunch through large volumes of data to generate actionable insights and intelligence. Enabling organizations to visualize & react to large amount of data in a clearer and more meaningful way.

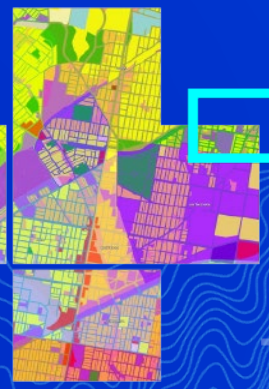
Helpful Links

- *What is GeoAnalytics Server?*
- *Detecting Incidents with Ozone Data*
- *Water Leaks Blog*
- *When to use GeoAnalytics Desktop Tools vs GeoAnalytics Server Tools Blog*
- *Data in GeoAnalytics Server*
- *Creating Heat Maps from Tracks Blog*
- *Dev Summit 2020 Plenary (Custom Spark Analysis) Blog and Video*
- *Dev Summit 2019 Plenary (Detecting Bus Delays) Blog and Video*

Email your questions at geoanalytics-pes@esri.com

esri

esri | THE SCIENCE OF WHERE



SEE WHAT OTHERS CAN'T