ArcGIS and Big Data: ArcGIS GeoAnalytics and the Spatiotemporal Big Data Store

Thomas Paschke
Agenda

1. What is “GeoAnalytics”?
2. Analysis Tools
3. Data Integration
4. Spatiotemporal Big Data Store
5. GeoAnalytics Desktop
6. Summary & Road Ahead
What is “GeoAnalytics”?
Every solution starts with a problem....

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.
ArcGIS Enterprise
with real-time & big data GIS capabilities

10.7

IoT

GeoEvent Server

spatiotemporal big data store

GeoAnalytics Server

Big Data

visualize

ingest, analyze

store

analyze
Analysis Tools
Analysis Capabilities
ArcGIS GeoAnalytics Server

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

**Use Proximity**
- Create Buffers

**Data Enrichment**
- Enrich from Multi-Variable Grid +

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

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

**Find Locations**
- Detect Incidents
- Find Similar Locations
- Geocode Locations

+ New at 10.7
Aggregation
GeoAnalytics Analysis Capabilities

- Aggregate points into polygons

- Aggregate points into bins

- Perform statistics on attributes of aggregation features (count, min, max, mean etc.)
Aggregation
GeoAnalytics Analysis Capabilities

- Aggregation into time steps
Time Stepping

GeoAnalytics Analysis Capabilities

• Three parameters to define a time step:
  - **Interval** (duration of time in a step)
  - **Repeat** (frequency of a step)
  - **Reference time** (alignment)

• Examples:
  - Hourly steps **Interval**: 1 hour
  - Every 12th hour **Interval**: 1 hour **Repeat**: 12 hours
  - Every Monday **Interval**: 1 day **Repeat**: 1 week **Reference**: Some Monday
Join Feature
GeoAnalytics Analysis Capabilities

Target Features  Join Features  Intermediate Result  Final Result

NOW  THEN
X  Y

NOW  THEN
X  Y
Join Feature
GeoAnalytics Analysis Capabilities

Operators in GeoAnalytics Server

**Spatial**
- Intersects
- Equals
- Near *
- Contains
- Within
- Touches
- Crosses
- Overlaps

**Temporal**
- Meets
- Met by
- Overlaps
- Overlapped by
- During
- Contains
- Equals
- Finishes
- Finished by
- Starts
- Started by
- Intersects
- Near *

**Attribute**
Features are joined based on a common attribute.

* Spatial and Temporal Near operator, allows you to specify a distance / time period:
Track Analysis
GeoAnalytics Analysis Capabilities

- Reconstruct Tracks

- Detect Incidents

The Detect Incidents tool examines time-sequential features using a specified condition. Features that meet the specified condition are marked as incidents.
Calculate Field
GeoAnalytics Analysis Capabilities

Calculates values for a new or existing field and creates a layer in your contents in ArcGIS Enterprise
Arcade in GeoAnalytics

GeoAnalytics Analysis Capabilities

- **Calculate Field** | Use to calculate new values. Can be time aware.
  - Calculate the values as the mean of the 3 previous values:
    \[
    \text{mean(TrackFieldWindow(“speed”, -4, -1))}
    \]

- **Reconstruct Tracks** | Optionally used to calculate the buffer distance. Can be time aware.
  - Buffer by $feature[“wake”]

- **Detect Incidents** | Used to detect features that meet a certain condition. Can be time aware.
  - Detect features that meet the condition:
    \[
    \text{abs($feature[“temperature”] - TrackFieldWindow(“temperature”, -1, 0)[0])} > 10
    \]
Demo: Arcade for Incident Detection
• **Append Data** | Optionally used to determine how values are appended
  - Append the mean($feature[“2015_Sales”], ($feature[“2016_Sales”], ($feature[“2017_Sales”])) to a field **AverageSales**

• **Join Features** | Used to specify which features should be used in a join.
  - Join Features that meet the condition $target[“cost”] > $join[“annual_cost”]

• **Create Buffers** | Optionally used to calculate the buffer distance.
  - Buffer by $feature[“Blast_Radius”] * 10
The **Find Hot Spots** tool will determine if there is any statistically significant clustering in the spatial pattern of your data.

The **Find Point Clusters** tool finds clusters of point features in surrounding noise based on their spatial distribution.

The **Forest-based Classification and Regression** tool models and generates predictions using an adaptation of Leo Breiman's random forest algorithm, which is a supervised machine learning method.

The **Generalized Linear Regression** tool generate predictions or models a dependent variable in terms of its relationship to a set of explanatory variables. This tool can be used to fit continuous (OLS), binary (logistic) and count (Poisson) models.
Demo: Detect Hot Spots of Delay
Data Enrichment
GeoAnalytics Analysis Capabilities

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

Enrich from Build Multi-Variable Grid
Access and use PySpark with GeoAnalytics Server

GeoAnalytics Analysis Capabilities

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:
Run a Python Script Tool
Data Integration
Data Integration
GeoAnalytics Server

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
ArcGIS Enterprise

with real-time & big data GIS capabilities

IoT

GeoEvent Server

spatiotemporal big data store

GeoAnalytics Server

10.7

big data file shares

feature service
Data Integration
Big Data File Shares

Input: Big Data File Share (Referenced data source)

• Direct read from:
  • File shares (local or network directories)
  • HDFS – Hadoop Distributed File System
  • Hive
  • Cloud Stores
    • AWS S3
    • Microsoft Azure Blob container
    • Microsoft Azure Data Lake

• Supported formats:
  • Delimited files (.csv, .tsv, .txt)
  • Shapefiles
  • Parquet files
  • ORC files

Output:

Hosted Feature Layer (SBDS)

Big Data File Share + (Referenced data source)
How do I actually use all those data sources?
4 Spatiotemporal Big Data Store
ArcGIS Enterprise
with real-time & big data GIS capabilities
The spatiotemporal big data store is a distributed, highly available data store

- Visualize on-the-fly aggregations of data
- Switch visualization from aggregation to raw features
- Perform exploratory queries over any combination of space, time, and attributes
Demo: Visualizing Big Data
spatiotemporal big data store

shards & replication factor
spatiotemporal big data store

auto-rebalancing of data upon node membership changes, + or -, in the big data store
spatiotemporal big data store

data retention policies, configured per data source
spatiotemporal big data store

automatic data backups using periodic snapshots, including ability to restore from a snapshot

GeoEvent Server

node 1

node 2

node 3

node 4

node 5

r = 1

GeoAnalytics Server

purge layer based on data retention policy

snapshot: 2016-05-17-11-0-0.snapshot
snapshot: 2016-05-17-12-0-0.snapshot
...

spatiotemporal big data store

manual and programmatic export data to cloud stores

GeoEvent Server

purge layer based on data retention policy

r = 1

node 1

node 5

node 2

node 4

node 3

GeoAnalytics Server

delimited text export
Azure Blob
Amazon S3

snapshot: 2016-05-17-11-0-0.snapshot
snapshot: 2016-05-17-12-0-0.snapshot
...
GeoAnalytics Desktop
GeoAnalytics is available through...

GeoAnalytics Server (10.5+) – use 1 or 3 machines to distribute analysis

ArcGIS Pro (2.4) – GeoAnalytics Desktop Tools – use your desktop machine
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

<table>
<thead>
<tr>
<th>Input data</th>
<th>GeoAnalytics Server (10.7 + 10.7.1)</th>
<th>GeoAnalytics Desktop (Pro 2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Big data file shares *</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Hosted feature layers</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Feature services</td>
<td>- Shapefiles</td>
</tr>
</tbody>
</table>

[Blog post covering this topic]
When to use Desktop or Server for GeoAnalytics

<table>
<thead>
<tr>
<th>Input data</th>
<th>GeoAnalytics Server (10.7 + 10.7.1)</th>
<th>GeoAnalytics Desktop (Pro 2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Big data file shares *</td>
<td></td>
<td>- File geodatabase</td>
</tr>
<tr>
<td>- Hosted feature layers</td>
<td></td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td>- Feature services</td>
<td></td>
<td>- Shapefiles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output data</th>
<th>GeoAnalytics Server (10.7 + 10.7.1)</th>
<th>GeoAnalytics Desktop (Pro 2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Hosted feature layers</td>
<td></td>
<td>- File geodatabase</td>
</tr>
<tr>
<td>- Big data file shares *</td>
<td></td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Shapefiles</td>
</tr>
</tbody>
</table>
When to use Desktop or Server for GeoAnalytics

<table>
<thead>
<tr>
<th></th>
<th>GeoAnalytics Server (10.7 + 10.7.1)</th>
<th>GeoAnalytics Desktop (Pro 2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input data</td>
<td>- Big data file shares *</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Hosted feature layers</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Feature services</td>
<td>- Shapefiles</td>
</tr>
<tr>
<td>Output data</td>
<td>- Hosted feature layers</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Big data file shares *</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td>Scaling out analysis</td>
<td>- Control the number of machines</td>
<td>- One machine only</td>
</tr>
<tr>
<td></td>
<td>- Control the percentage of cores and RAM</td>
<td>- Control the percentage of RAM (or a value)</td>
</tr>
<tr>
<td></td>
<td>- Scale out data storage with spatiotemporal data store</td>
<td></td>
</tr>
</tbody>
</table>
## When to use Desktop or Server for GeoAnalytics

<table>
<thead>
<tr>
<th></th>
<th>GeoAnalytics Server (10.7 + 10.7.1)</th>
<th>GeoAnalytics Desktop (Pro 2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input data</strong></td>
<td>- Big data file shares *</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Hosted feature layers</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Feature services</td>
<td>- Shapefiles</td>
</tr>
<tr>
<td><strong>Output data</strong></td>
<td>- Hosted feature layers</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Big data file shares *</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Shapefiles</td>
</tr>
<tr>
<td><strong>Scaling out analysis</strong></td>
<td>- Control the number of machines</td>
<td>- One machine only</td>
</tr>
<tr>
<td></td>
<td>- Control the percentage of cores and RAM</td>
<td>- Control the percentage of RAM</td>
</tr>
<tr>
<td></td>
<td>- Scale out data storage with spatiotemporal data store</td>
<td>(or a value)</td>
</tr>
<tr>
<td><strong>Tools</strong></td>
<td>- 26 tools (and adding more)</td>
<td>- 15 tools (and adding more)</td>
</tr>
<tr>
<td></td>
<td>- Run Python Script</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GeoAnalytics Server (10.7 + 10.7.1)</td>
<td>GeoAnalytics Desktop (Pro 2.4)</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>Input data</strong></td>
<td>- Big data file shares *</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Hosted feature layers</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Feature services</td>
<td>- Shapefiles</td>
</tr>
<tr>
<td><strong>Output data</strong></td>
<td>- Hosted feature layers</td>
<td>- File geodatabase</td>
</tr>
<tr>
<td></td>
<td>- Big data file shares *</td>
<td>- Enterprise geodatabase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Shapefiles</td>
</tr>
<tr>
<td><strong>Scaling out analysis</strong></td>
<td>- Control the number of machines</td>
<td>- One machine only</td>
</tr>
<tr>
<td></td>
<td>- Control the percentage of cores and RAM</td>
<td>- Control the percentage of RAM (or a value)</td>
</tr>
<tr>
<td></td>
<td>- Scale out data storage with spatiotemporal data store</td>
<td></td>
</tr>
<tr>
<td><strong>Tools</strong></td>
<td>- 26 tools (and adding more)</td>
<td>- 15 tools (and adding more)</td>
</tr>
<tr>
<td></td>
<td>- Run Python Script</td>
<td></td>
</tr>
<tr>
<td><strong>Interface</strong></td>
<td>- REST + the ArcGIS API for Python</td>
<td>- Pro and Arcpy (and model builder)</td>
</tr>
<tr>
<td></td>
<td>- Pro and Arcpy (and model builder)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Portal Map Viewer</td>
<td></td>
</tr>
</tbody>
</table>
Demo: GeoAnalytics Desktop Tools
Summary and Road Ahead
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.
Road Ahead

• *Continued performance improvements (and bug fixes)* – do you have any to report?!

• *Analysis*
  • Find dwell location
  • Geographically weighted regression
  • Added spatiotemporal clustering to Find Point Clusters
  • Track Analysis

• *Data Management:*
  • Adding more data sources.
  • Working towards adding big data file share like experience in Pro.
Please Take Our Survey on the App

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the feedback section

Complete answers and select “Submit”