Esri Developer Summit

March 8–11, 2016 | Palm Springs, CA

BIG DATA AND ARCGIS

Michael Park

Erik Hoel

ArcGIS 2016

Feature Analytics and Big Data Processing: What's Coming in 2016

What is Feature Geo Analytics?

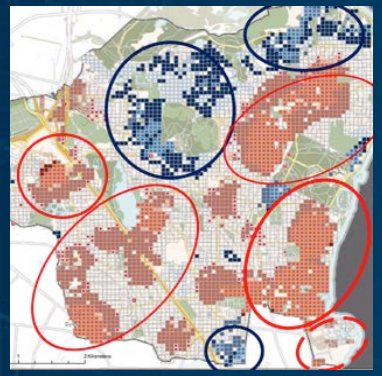
*ArcGIS has a new way of processing **spatiotemporal data** that is designed for **big data** by leveraging **distributed analytics** and storage*

- **Works with your existing GIS data and tabular data**
- **Designed to perform both spatial and temporal analysis**
- **Uses familiar workflows to complete complex analyses**

Feature Geo Analytics

- Feature Geo Analytics adds to existing ArcGIS analysis capabilities

Geoprocessing



Feature Geo Analytics

Web GIS Layers

Powerful analytics

more

Distributed analytics and storage

new

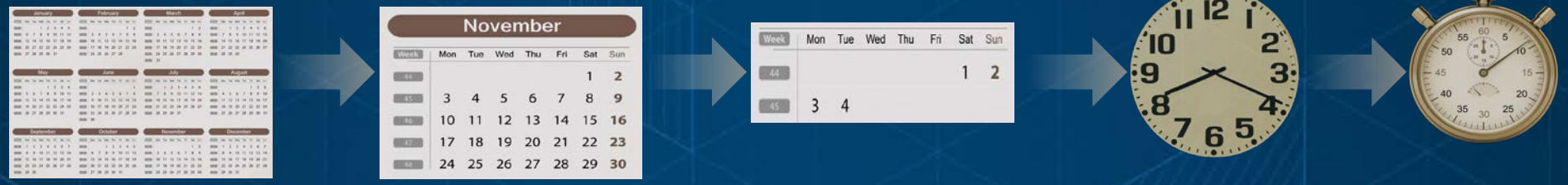
Rich geoinformation model

extends

Solve New Problems

Run analytics:

- **against data that is too big for a single desktop machine**
  - hundreds of millions of 911 calls accumulated over years
  - billions of observations of ship movements ingested through GeoEvent
- **designed to gain insight into both spatial and temporal patterns**
- **against massive collections in a scalable manner**
- **and meet time constraints**



months weeks days hours minutes

Rich Collection of Analysis Tools

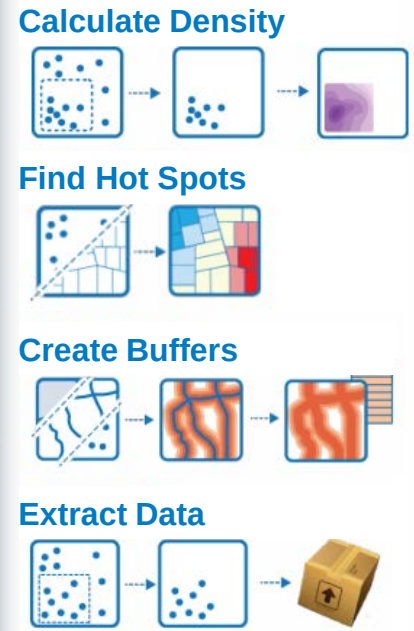
Summarize Data
- Aggregate Points
- Summarize Nearby
- Summarize Within
- Reconstruct Tracks
- Join Features

Find Locations
- Find Existing Locations
- Find Similar Locations

Analyze Patterns
- Calculate Density
- Find Hot Spots
- Create Space Time Cube

Use Proximity
- Create Buffers

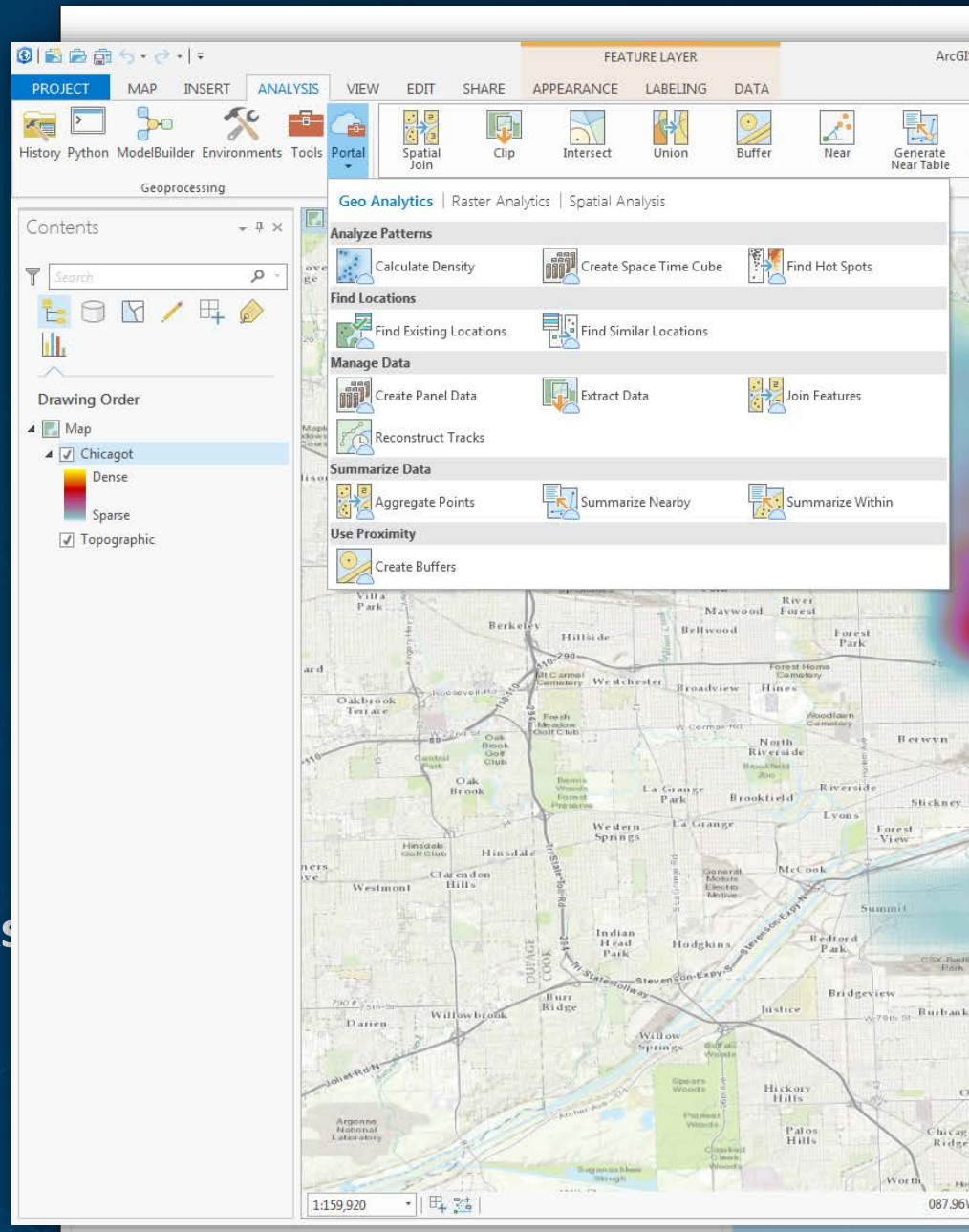
Manage Data
- Extract Data



** Temporally aware tools*

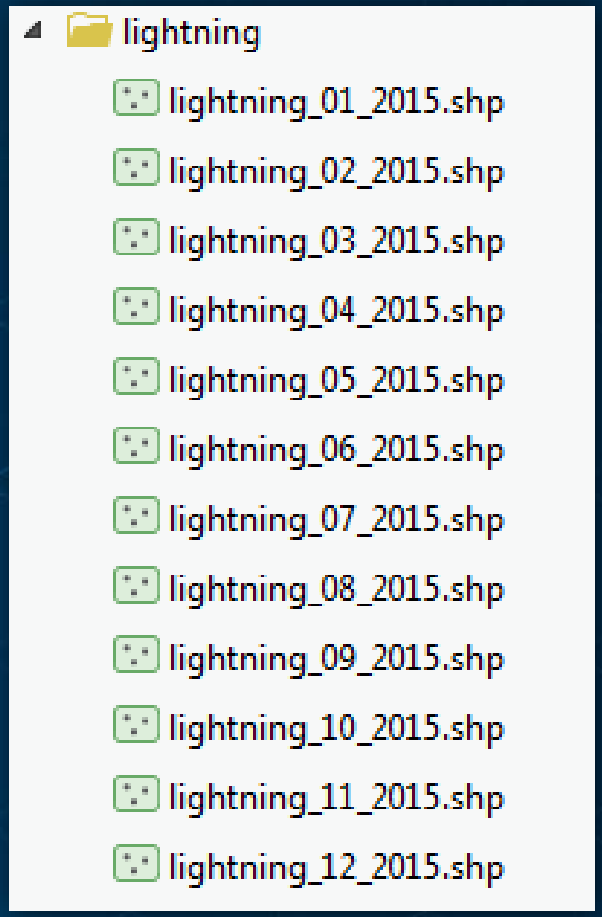
Familiar and Simple Interfaces

- **Easy to get started, available "out of the box"**
- **Simple user experience**
  - Simplified analysis Web UX through **Portal**
  - Familiar tool box and ribbon experience through **Pro**
- **Great developer interfaces**
  - Python developer experience **through Pro**
  - Develop quickly and share through **Python notebooks**
  - Familiar **REST API**



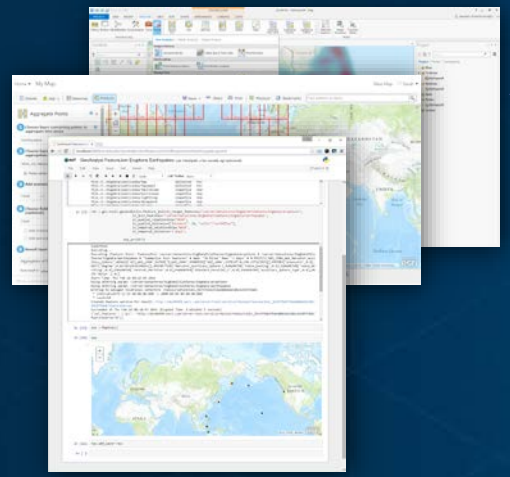
Feature Geo Analytics and Your Data

- **Use your own big data sources**
  - Easily attach folders located locally on a disk or a file share
    - Multiple files can be recognized as a single Geo Analytics dataset
  - Attach big data stores
- **Use your GIS data**
  - Works with layers already in your Web GIS



Geo Analytics Architectural Overview

Feature Geo Analytics can power systems that need to execute analysis in a distributed and scalable environment. It is designed for analysts, researchers, and developers.



Use your Web GIS layers through Pro, Portal, Python Notebooks, or the REST API

New Web GIS Layers

Web GIS Layers

Portal

Server Cluster

Shapefiles

Delimited Files

Enterprise

Big Data Stores

Relational Data Store

Spatiotemporal Data Store

Un-Managed Data

Managed Data

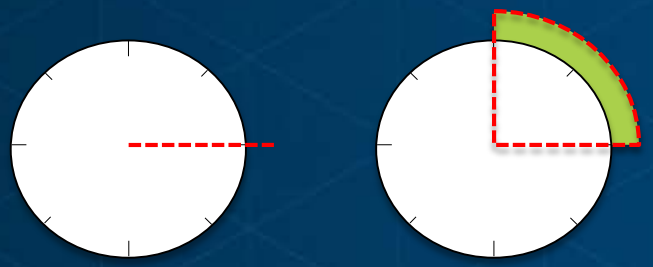
ArcGIS 2016

Feature Geo Analytics: Analytic Details

Analytical Overview

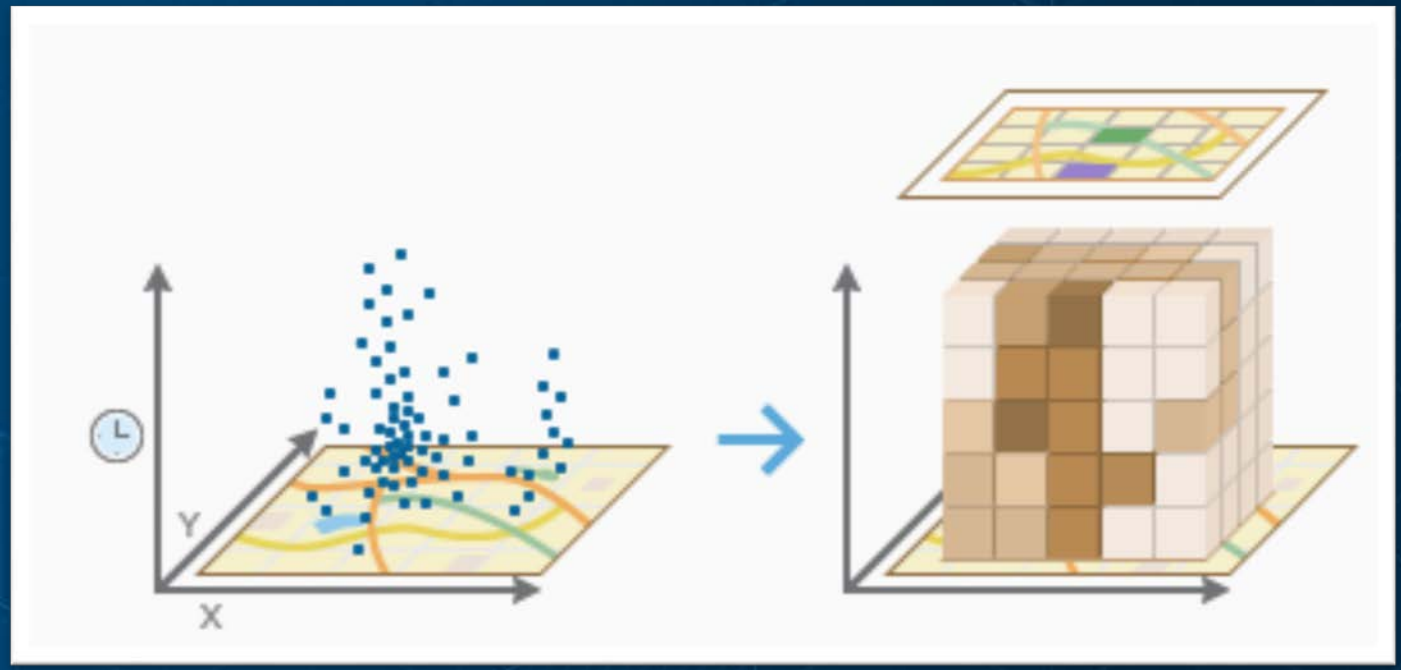
- Analysis capabilities patterned from ArcGIS Online Spatial Analysis

- Tools useful for big data with an emphasis on:
  - Spatiotemporal analysis
  - Summarization and aggregation



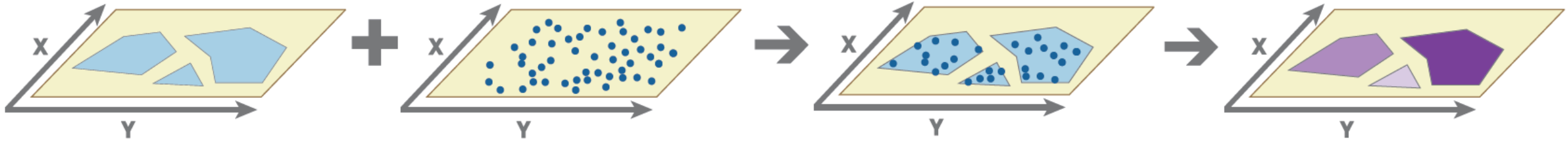
Instant

Interval

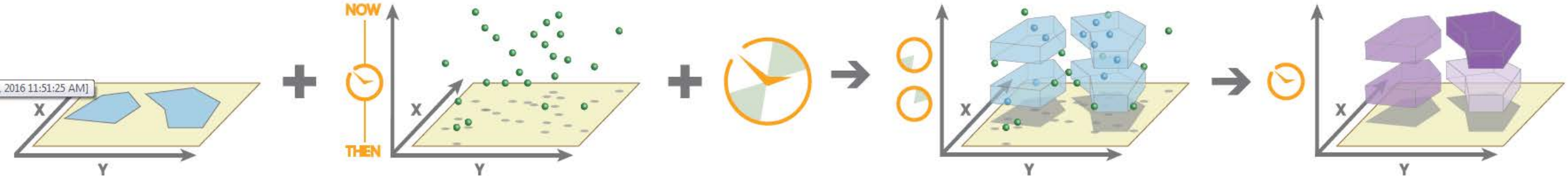


Aggregation and Summarization

- Spatial Joins

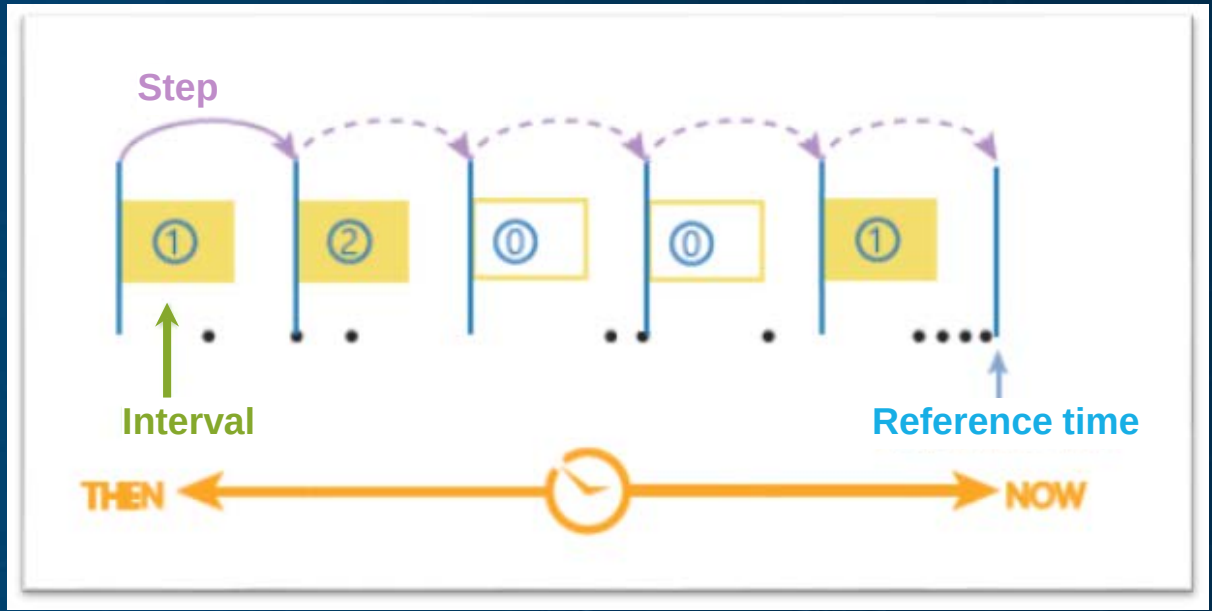


- Space-time slices



Time Slicing

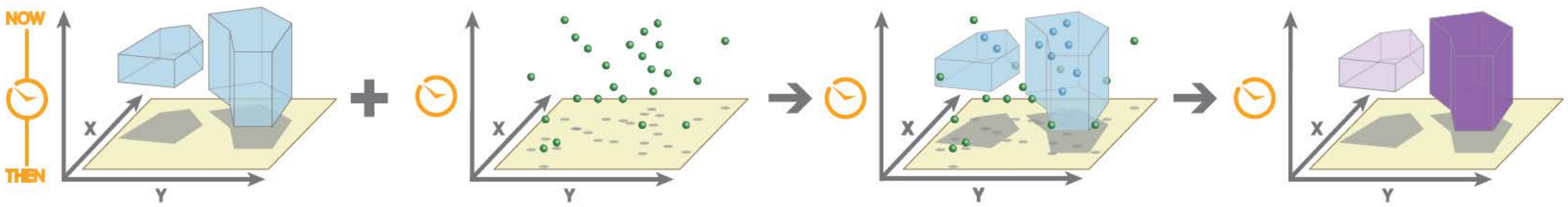
- Three parameters to define a time slice:
  - Interval (time in a slice)
  - Step (frequency of slice)
  - Reference time (alignment)



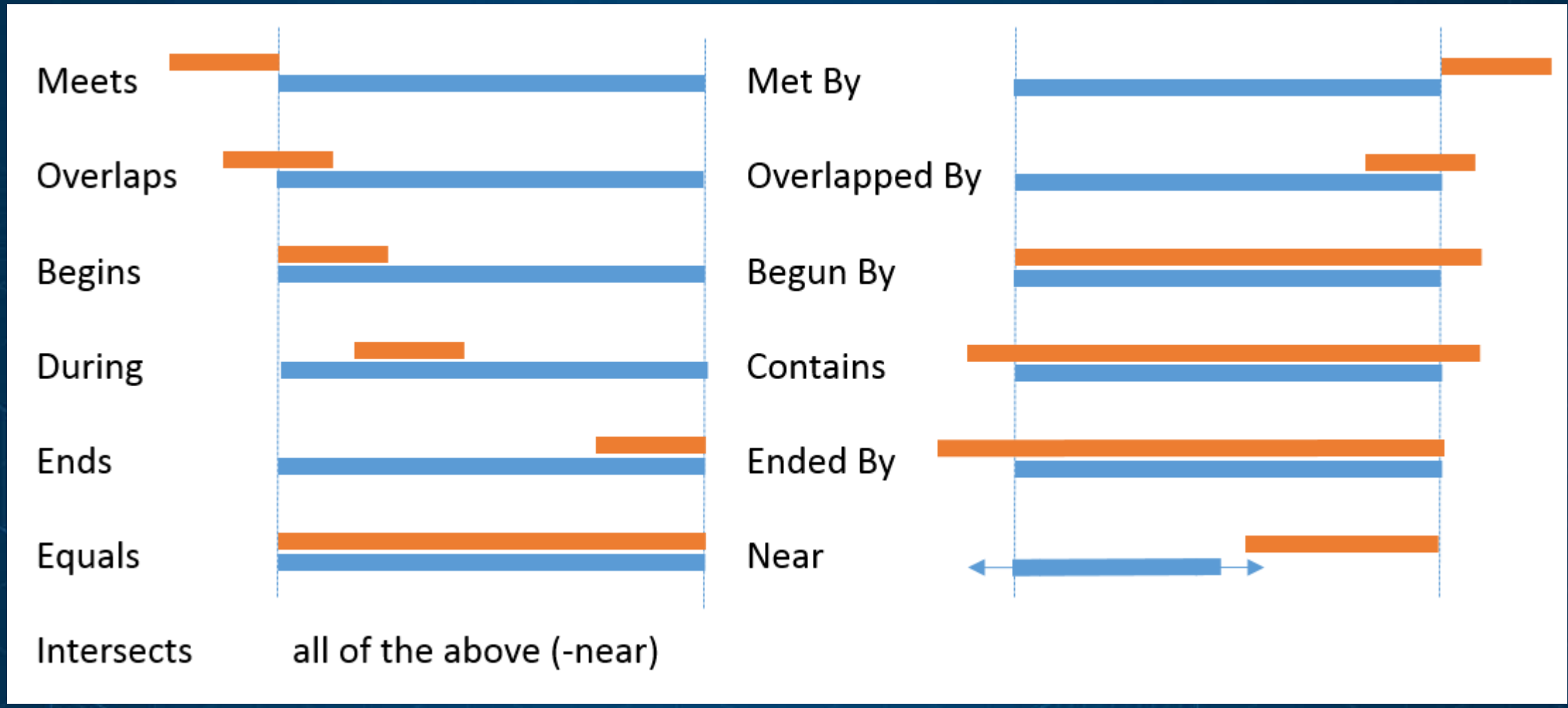
- Examples:
  - Hourly slices — Interval: 1 hour
  - Every 12th hour — Interval: 1 hour — Step: 12 hours
  - Every Monday — Interval: 1 day — Step: 1 week — Reference: Some Monday

Aggregation and Summarization (cont.)

- **Spatiotemporal joins**

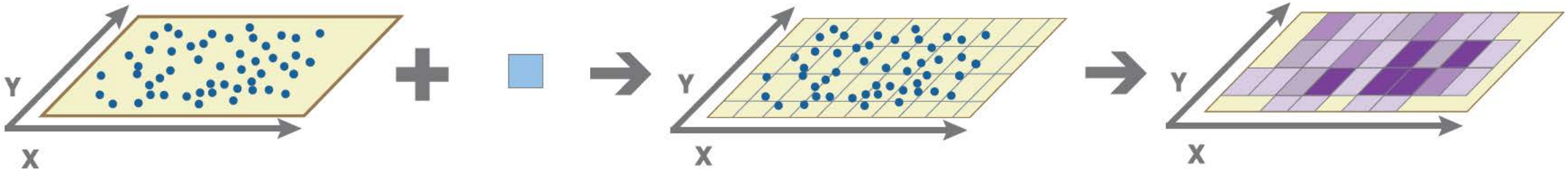


Temporal Relationships on Intervals



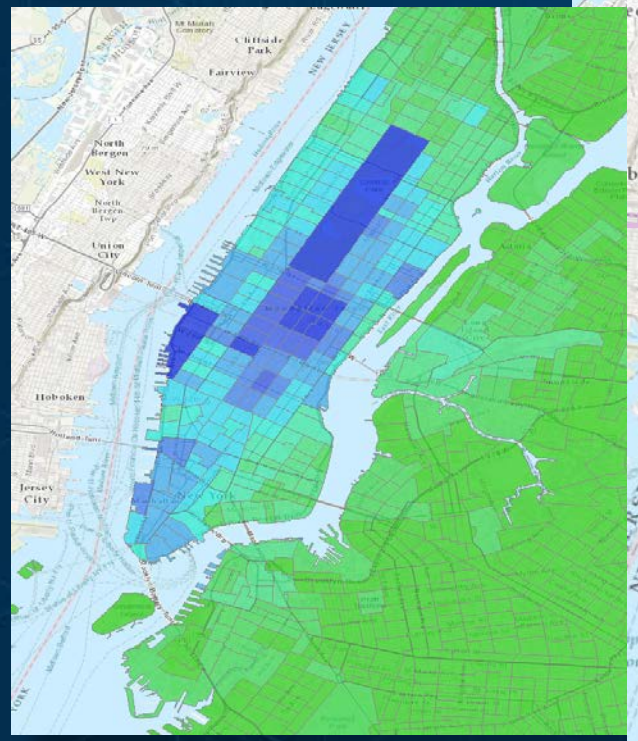
Aggregation and Summarization (cont.)

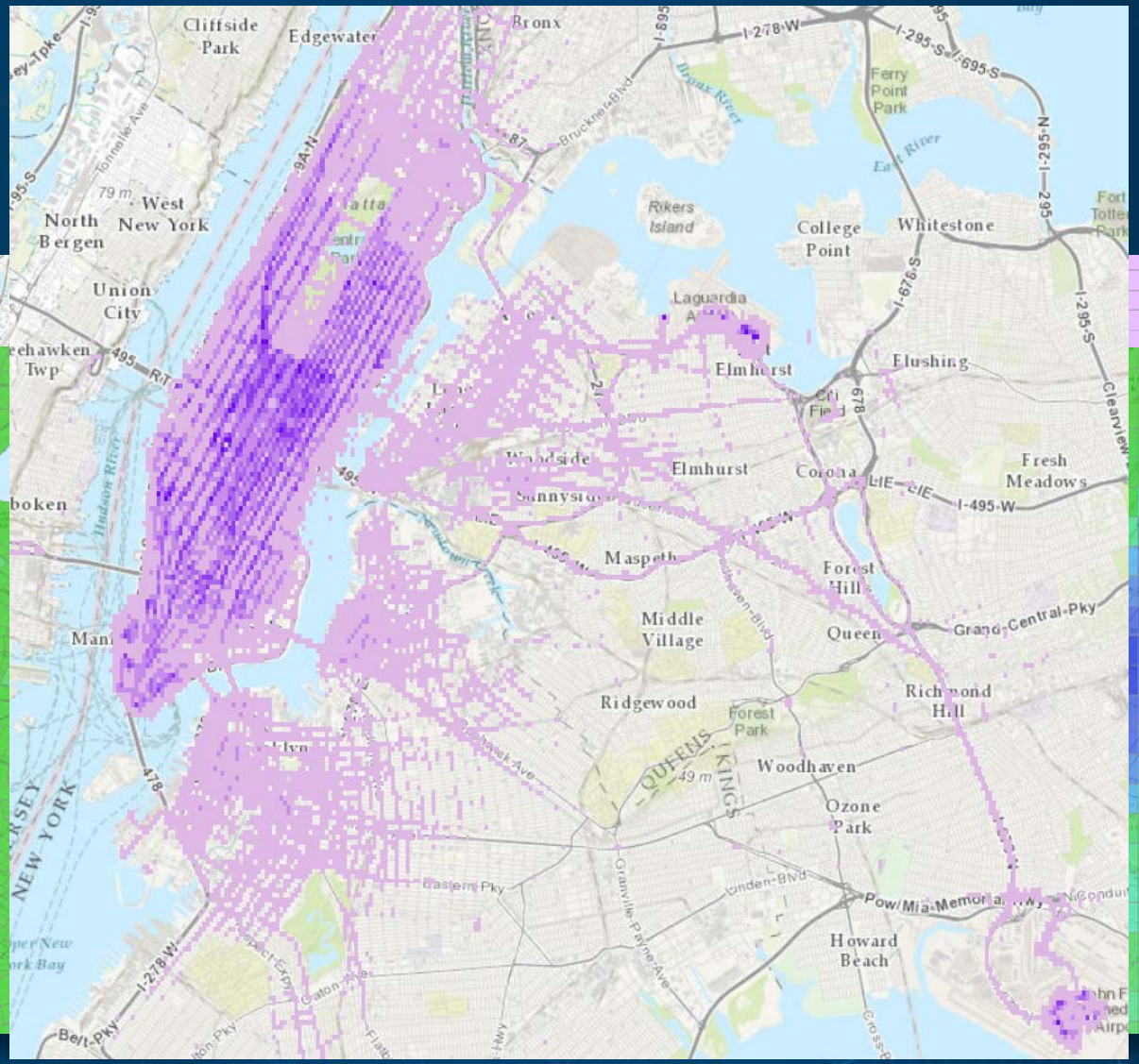
- Points into Bins



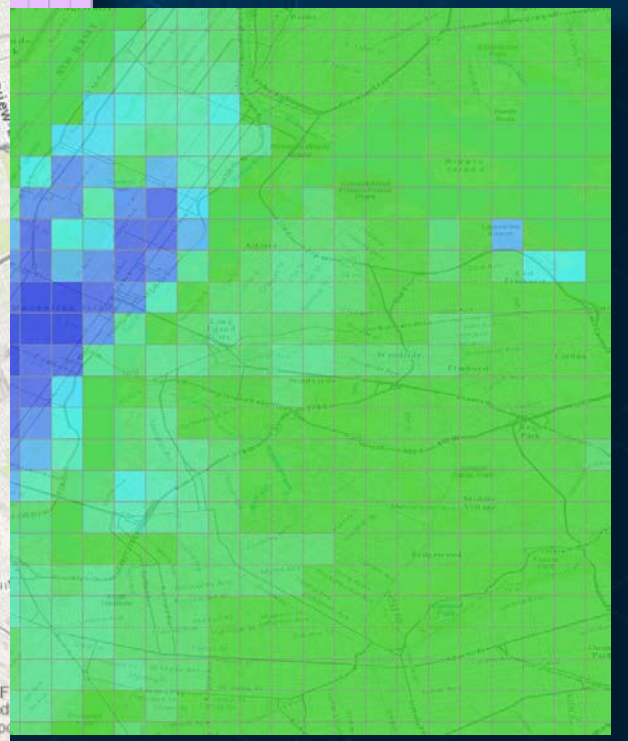
Aggregation – Polygons vs Cells

Aggregation By Polygons



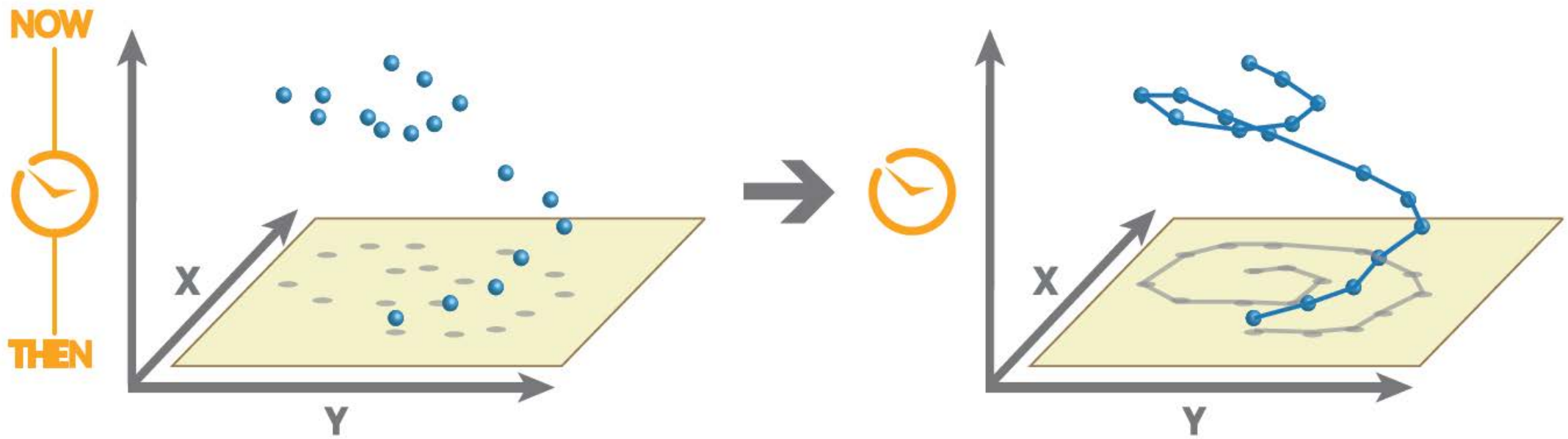


Aggregation By Cells



Aggregation and Summarization (cont.)

- Reconstruct Tracks
  - Summarize time enabled points into tracks

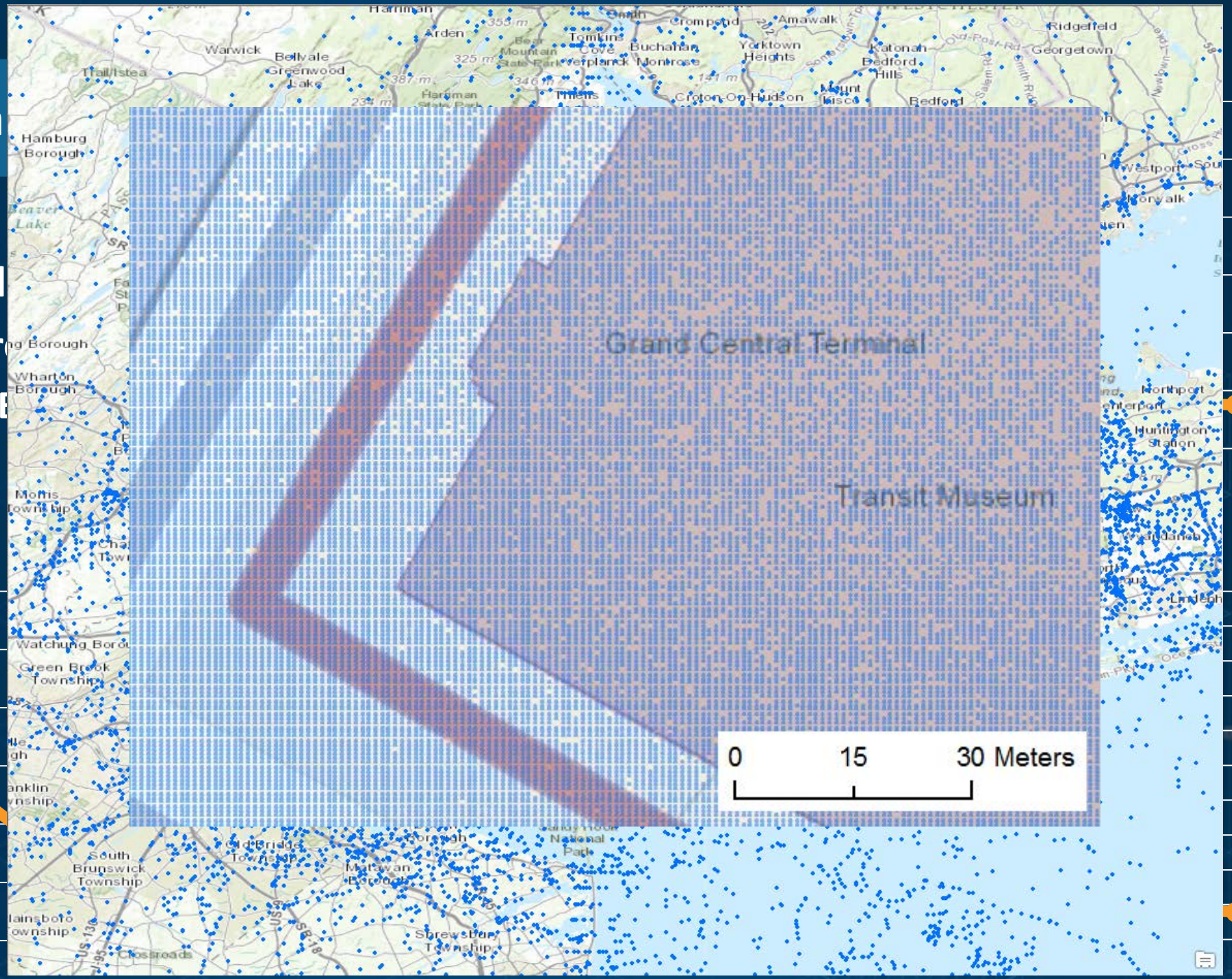


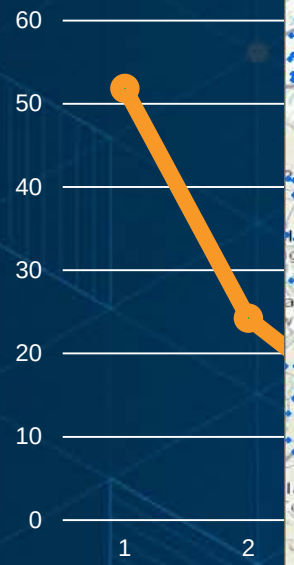
Demo – Construct Tracts

- Hurricane dataset
  - 120,000 points, ~100 years
  - Each point has:
    - Serial number
    - Location
    - Date
    - Wind speed and pressure attributes
  - Problems?
    - Difficult to visualize that many points
    - Difficult to visualize hurricane path

Geo Ana

- 16 nod
  - 4 cor
  - 14 GB





Geo Analytics Performance

- 16 nodes in the cluster
  - 4 cores per node (commodity desktops)
  - 14 GB RAM per node

- Aggregate 4 billion points into polygons
  - 52 polygons (states) — 12 minutes
  - 11 million polygons (census blocks) — 1.2 hours

4 billion points x 11 million polygons

200K x more polygons – 6 x more time

Geo Analytics Performance

- 18 nodes in the cluster
  - 8 cores per node (commodity desktops)
  - 24 GB RAM per node

- Spatiotemporal attribute join of 1 billion points (the plenary financial dataset)
  - 1 km radius
  - 1 hour time window
  - target:destination_id = join:destination_id (i.e., transactions to the same bank)

- **<1 hour, including all I/O to ArcGIS Spatiotemporal Data Store**

Why will I want to use Geo Analytics?

- Helps you get “big jobs” done faster
- Makes big data easier to use and easier to share
- Parallelizes and focuses on space and time analysis for massive dataset processing
- Feature Geo Analytics are “out of the box” and “ready to use” within your ArcGIS system
  - familiar and simple interfaces
  - faster prototyping, R&D, and insight into your data
- Lets you see the big data picture in a clearer and more meaningful way

esri®