Real-Time & Big Data GIS: Best Practices

Josh Joyner
Adam Mollenkopf
ArcGIS Enterprise

with real-time capabilities

Apps
Desktop
APIs
ArcGIS Enterprise
ArcGIS GeoEvent Server
spatiotemporal big data store
analytics
storage
visualizations
live & historic aggregates & features
live features stream services
ingestion
actuation
Agenda:

1. Architecture Recommendations
2. Big Data Storage
3. Performance, Resiliency, & Scalability
4. Stream Services
5. Service Design Considerations
6. Upgrade Planning
7. Troubleshooting
GeoEvent Server

What are the primary factors I should consider?

- **Operating environment:**
  - virtual machines – beware! resources need to be shared in an effective way, like EC2 or Azure.
  - dedicated bare metal machines or public cloud instances are much more deterministic.

- **Network**
  - speed – the faster the better.  

- **Memory**
  - size – 8GB has been required since 10.3. 
  - type – minimum of DDR3 is recommended.
  - clock speed (MHz) and transfer rate (Mbps) – the faster the better.

- **Processors**
  - # of cores – the more the better.

- **Disk**
  - 700MB required for installation
  - amount of disk space needed will vary based on quantity of deployed input connectors
  - each input can utilize up to a maximum of 600 MB of disk space before clean up

**m5.2xlarge**

1 GBit/s

32GiB, *default JVM max heap size is 4 GB*

8 vCPU

10GB recommended minimum  *(new for 10.6)*
spatiotemporal big data store

What are the primary factors I should consider?

- **Operating environment:**  
  - virtual machines – beware! resources need to be shared in an effective way, like EC2 or Azure.  
  - dedicated bare metal machines or public cloud instances are much more deterministic.

- **Disk**  
  - speed – the faster the better  
  - 1,000 Mbps EBS, note: local SSD is much better

- **Network**  
  - speed – the faster the better.  
  - 1 GBit/s

- **Memory**  
  - size – 16GB minimum.  
  - type – DDR3 is recommended.  
  - clock speed (MHz) and transfer rate (Mbps) – the faster the better.  

 32GiB, big data store allocates 8GiB by default

- **Processors**  
  - # of cores – the more the better.  
  - speed (GHz) – the faster the better.

 8 vCPU
ArcGIS Enterprise

*with real-time capabilities*

**MINIMUM environment**

3 machines

---

**IoT**

**ArcGIS Enterprise**

visualize

**GeoEvent Server**

 ingest, analyze

**spatiotemporal big data store**

**store**

---

**functional servers & spatiotemporal big data store**

**SHOULD BE on ISOLATED machines**
ArcGIS Enterprise

with real-time capabilities

functional servers & spatiotemporal big data store

SHOULD BE on ISOLATED machines
ArcGIS Enterprise
with real-time & big data GIS capabilities
ArcGIS Enterprise

*with real-time & big data GIS capabilities*

**OPTIMIZED environment**
10 machines

---

1. ArcGIS Enterprise
2. GeoEvent Server
3. IoT
4. Ingest, analyze
5. Spatiotemporal big data store
6. Store
7. GeoAnalytics Server
8. Analyze
9. Big Data
10. Green arrows
ArcGIS Enterprise
with real-time & big data GIS capabilities on Microsoft Azure

1. **Microsoft Azure**
   - ArcGIS Enterprise Cloud Builder for Microsoft Azure

2. **GeoEvent Server**
   - Ingest, analyze

3. **GeoEvent Server**
   - GeoEvent Server

4. **GeoEvent Server**
   - GeoEvent Server

5. **spatiotemporal big data store**
   - Store

6. **spatiotemporal big data store**
   - spatiotemporal big data store

7. **GeoAnalytics Server**
   - Analyze

8. **GeoAnalytics Server**
   - GeoAnalytics Server

9. **GeoAnalytics Server**
   - GeoAnalytics Server

10. **Big Data**
    - Big Data

**ArcGIS Enterprise 10.6.1**
ArcGIS Enterprise
with real-time & big data GIS capabilities on Amazon EC2

ArcGIS Enterprise
Cloud Builder
for Amazon EC2

Big Data

GeoEvent Server
 ingest, analyze

spatiotemporal big data store
 store

GeoAnalytics Server
 analyze
2 Big Data Storage
ArcGIS Enterprise

with real-time capabilities

OPTIMIZED environment
5 machines

1. ArcGIS Enterprise

2. GeoEvent Server

3. spatiotemporal big data store

IoT

visualize

ingest, analyze

store
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

GeoEvent Server

node 1

node 2

node 3

node 4

node 5

r = 1

purge based on data retention
spatiotemporal big data store

*rolling index option, set appropriately to the velocity of your observation data*
spatiotemporal big data store

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

**choosing an Object Id option**
<table>
<thead>
<tr>
<th>IoT rate per second</th>
<th>per day</th>
<th>per week</th>
<th>per month</th>
<th>per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86,400</td>
<td>604,800</td>
<td>2,592,000</td>
<td>31,104,000</td>
</tr>
<tr>
<td>10</td>
<td>864,000</td>
<td>6,048,000</td>
<td>25,920,000</td>
<td>311,040,000</td>
</tr>
<tr>
<td>100</td>
<td>8,640,000</td>
<td>60,480,000</td>
<td>259,200,000</td>
<td>3,110,400,000</td>
</tr>
<tr>
<td>1,000</td>
<td>86,400,000</td>
<td>604,800,000</td>
<td>2,592,000,000</td>
<td>31,104,000,000</td>
</tr>
<tr>
<td>10,000</td>
<td>864,000,000</td>
<td>6,048,000,000</td>
<td>25,920,000,000</td>
<td>311,040,000,000</td>
</tr>
<tr>
<td>100,000</td>
<td>8,640,000,000</td>
<td>60,480,000,000</td>
<td>259,200,000,000</td>
<td>3,110,400,000,000</td>
</tr>
<tr>
<td>1,000,000</td>
<td>86,400,000,000</td>
<td>604,800,000,000</td>
<td>2,592,000,000,000</td>
<td>31,104,000,000,000</td>
</tr>
</tbody>
</table>

Choosing an Object Id option

Max Value | # of IDs | ArcGIS Clients
---|---------|-----------------
Int32 | 2,147,483,647 | 2.1 billion | Pro, Desktop, Ops Dashboard, …
### Max Value

| ArcGIS Clients | Int32         | 2,147,483,647 | 2.1 billion |

### IoT rate per second

<table>
<thead>
<tr>
<th>IoT rate per second</th>
<th>per day</th>
<th>per week</th>
<th>per month</th>
<th>per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86,400</td>
<td>604,800</td>
<td>2,592,000</td>
<td>31,104,000</td>
</tr>
<tr>
<td>10</td>
<td>864,000</td>
<td>6,048,000</td>
<td>25,920,000</td>
<td>311,040,000</td>
</tr>
<tr>
<td>100</td>
<td>8,640,000</td>
<td>60,480,000</td>
<td>259,200,000</td>
<td>3,110,400,000</td>
</tr>
<tr>
<td>1,000</td>
<td>86,400,000</td>
<td>604,800,000</td>
<td>2,592,000,000</td>
<td>31,104,000,000</td>
</tr>
<tr>
<td>10,000</td>
<td>864,000,000</td>
<td>6,048,000,000</td>
<td>25,920,000,000</td>
<td>311,040,000,000</td>
</tr>
<tr>
<td>100,000</td>
<td>8,640,000,000</td>
<td>60,480,000,000</td>
<td>259,200,000,000</td>
<td>3,110,400,000,000</td>
</tr>
<tr>
<td>1,000,000</td>
<td>86,400,000,000</td>
<td>604,800,000,000</td>
<td>2,592,000,000,000</td>
<td>31,104,000,000,000</td>
</tr>
</tbody>
</table>
### Spatiotemporal Big Data Store

**Choosing an Object Id Option**

<table>
<thead>
<tr>
<th></th>
<th>Max Value</th>
<th># of IDs</th>
<th>ArcGIS Clients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int32</td>
<td>2,147,483,647</td>
<td>2.1 billion</td>
<td>Pro, Desktop, Ops Dashboard, …</td>
</tr>
<tr>
<td>Int64 (signed)</td>
<td>9,223,372,036,854,775,807</td>
<td>9.2 quintillion</td>
<td>JavaScript, custom apps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>IoT rate per second</strong></th>
<th><strong>per day</strong></th>
<th><strong>per week</strong></th>
<th><strong>per month</strong></th>
<th><strong>per year</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86,400</td>
<td>604,800</td>
<td>2,592,000</td>
<td>31,104,000</td>
</tr>
<tr>
<td>10</td>
<td>864,000</td>
<td>6,048,000</td>
<td>25,920,000</td>
<td>311,040,000</td>
</tr>
<tr>
<td>100</td>
<td>8,640,000</td>
<td>60,480,000</td>
<td>259,200,000</td>
<td>3,110,400,000</td>
</tr>
<tr>
<td>1,000</td>
<td>86,400,000</td>
<td>604,800,000</td>
<td>2,592,000,000</td>
<td>31,104,000,000</td>
</tr>
<tr>
<td>10,000</td>
<td>864,000,000</td>
<td>6,048,000,000</td>
<td>25,920,000,000</td>
<td>311,040,000,000</td>
</tr>
<tr>
<td>100,000</td>
<td>8,640,000,000</td>
<td>60,480,000,000</td>
<td>259,200,000,000</td>
<td>3,110,400,000,000</td>
</tr>
<tr>
<td>1,000,000</td>
<td>86,400,000,000</td>
<td>604,800,000,000</td>
<td>2,592,000,000,000</td>
<td>31,104,000,000,000</td>
</tr>
<tr>
<td></td>
<td>Max Value</td>
<td># of IDs</td>
<td>ArcGIS Clients</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------</td>
<td>----------------</td>
<td>-----------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Int32</td>
<td>2,147,483,647</td>
<td>2.1 billion</td>
<td>Pro, Desktop, Ops Dashboard, …</td>
<td></td>
</tr>
<tr>
<td>Int64 (signed)</td>
<td>9,223,372,036,854,775,807</td>
<td>9.2 quintillion</td>
<td>JavaScript, custom apps</td>
<td></td>
</tr>
<tr>
<td>UniqueStringID</td>
<td>n/a</td>
<td>unlimited</td>
<td>JavaScript, custom apps</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IoT rate per second</th>
<th>per day</th>
<th>per week</th>
<th>per month</th>
<th>per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86,400</td>
<td>604,800</td>
<td>2,592,000</td>
<td>31,104,000</td>
</tr>
<tr>
<td>10</td>
<td>864,000</td>
<td>6,048,000</td>
<td>25,920,000</td>
<td>311,040,000</td>
</tr>
<tr>
<td>100</td>
<td>8,640,000</td>
<td>60,480,000</td>
<td>259,200,000</td>
<td>3,110,400,000</td>
</tr>
<tr>
<td>1,000</td>
<td>86,400,000</td>
<td>604,800,000</td>
<td>2,592,000,000</td>
<td>31,104,000,000</td>
</tr>
<tr>
<td>10,000</td>
<td>864,000,000</td>
<td>6,048,000,000</td>
<td>25,920,000,000</td>
<td>311,040,000,000</td>
</tr>
<tr>
<td>100,000</td>
<td>8,640,000,000</td>
<td>60,480,000,000</td>
<td>259,200,000,000</td>
<td>3,110,400,000,000</td>
</tr>
<tr>
<td>1,000,000</td>
<td>86,400,000,000</td>
<td>604,800,000,000</td>
<td>2,592,000,000,000</td>
<td>31,104,000,000,000</td>
</tr>
</tbody>
</table>
Performance, Resiliency, & Scalability

3
Performance, Resiliency & Scalability

throughput benchmarks

- Throughput has increased with every release of ArcGIS GeoEvent Server
  - 10.6 release can support up to 6,000 events per second (e/s)

<table>
<thead>
<tr>
<th>ArcGIS GeoEvent Server</th>
<th>10.2</th>
<th>10.3</th>
<th>10.4</th>
<th>10.5</th>
<th>10.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity throughput</td>
<td>up to</td>
<td>up to</td>
<td>up to</td>
<td>up to</td>
<td>up to</td>
</tr>
<tr>
<td>measured in events per second (e/s)</td>
<td>500 e/s</td>
<td>2,000 e/s</td>
<td>3,000 e/s</td>
<td>4,000 e/s</td>
<td>6,000 e/s</td>
</tr>
</tbody>
</table>
Performance, Resiliency & Scalability

throughput benchmarks

- Throughput has increased with every release of ArcGIS GeoEvent Server
  - 10.6 release can support up to 6,000 events per second (e/s)
  - 10.6.1 release can support up to 10,000 events per second (e/s)

<table>
<thead>
<tr>
<th>ArcGIS GeoEvent Server</th>
<th>10.2</th>
<th>10.3</th>
<th>10.4</th>
<th>10.5</th>
<th>10.6</th>
<th>10.6.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity throughput</td>
<td>up to 500 e/s</td>
<td>up to 2,000 e/s</td>
<td>up to 3,000 e/s</td>
<td>up to 4,000 e/s</td>
<td>up to 6,000 e/s</td>
<td>up to 10,000 e/s</td>
</tr>
</tbody>
</table>
Performance, Resiliency & Scalability

*multi-machine site support*

### ArcGIS 10.5
- Resiliency (high availability) & scalability is only possible if users “bring their own gateway”.
  - Barrier to entry is HIGH & typically requires a professional services engagement for success.
  - Loses flexibility of input types.

### ArcGIS 10.6
- Provides users with a resilient & scalable Real-Time GIS deployment **OUT-OF-THE-BOX**.
  - Introduces a gateway process that is automatically configured as part of GeoEvent Server installation.
  - Provides flexibility for all input types.

<table>
<thead>
<tr>
<th>ArcGIS GeoEvent Server</th>
<th>10.2</th>
<th>10.3</th>
<th>10.4</th>
<th>10.5</th>
<th>10.6</th>
<th>10.6.1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Velocity throughput</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>measured in events per second (e/s)</em></td>
<td>up to 500 e/s</td>
<td>up to 2,000 e/s</td>
<td>up to 3,000 e/s</td>
<td>up to 4,000 e/s</td>
<td>up to 6,000 e/s</td>
<td>up to 10,000 e/s</td>
</tr>
<tr>
<td><strong>Resiliency &amp; Scalability</strong></td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><em>via multi-machine site</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Performance, Resiliency & Scalability

*multi-machine site support*


### Tutorial - GeoEvent Server 10.6.x Multiple-Machine Site

**Overview**

ArcGIS GeoEvent Server 10.6.x now supports the creation of multiple-machine sites. Details

- Document Link by GeoEventTeam
- Created: Mar 3, 2018   Updated: Mar 3, 2018
- View Count: 0

**Description**

ArcGIS GeoEvent Server 10.6.x now supports the creation of multiple-machine sites. In a multiple-machine site, two or more GeoEvent Server machines can be administered and used as a single logical unit, providing GeoEvent Server administrators with great flexibility to easily adjust the computing power of the site by adding or removing GeoEvent Server machines.

This tutorial will walk you through how to plan, setup, and work with a GeoEvent Server 10.6.x multiple-machine site. Also included is an appendix for administrators to learn how to monitor an existing GeoEvent Server multiple-machine site.
Performance, Resiliency & Scalability

Factors that influence throughput

Input event counts don’t always tell the whole story
Performance, Resiliency & Scalability

configuration changes to support larger scale

- GeoEvent Server by default is only allocated 4GB of RAM for the JVM
  - If utilizing a large amount of GeoFences it may be necessary to increase this amount
- This can be modified through the “/etc/ArcGISGeoEvent.cfg” up to 32GB (JVM limitation)
Performance, Resiliency & Scalability
configuration changes to support larger scale

- GeoEvent Server by default is able to maintain the state of 1000 unique Track_IDs.
  - This value can be increased by editing "/etc/com.esri.ges.manager.servicemanager.cfg"

- You may also need to modify the Incident Manager Setting in the Global Setting Tab if used in conjunction with the Incident Detector Processor
Stream Services
10.5 best practice = isolated deployment of GeoEvent Server (site per GeoEvent)

- An isolated deployment of GeoEvent instances leads to challenges with Stream Services:
  - Client A & B see event 1, while client C & D do not
  - Client C & D see event 2, while client A & B do not
Stream Service resilience & scalability

10.5 best practice = isolated deployment of GeoEvent Server (site per GeoEvent)

- GeoEvent instances input configuration use separate consumer groups:
  - With this configuration, all clients see all events

![Diagram showing GeoEvent server instances and stream services](image)
Stream Service resilience & scalability

10.5 best practice = isolated deployment of GeoEvent Server (site per GeoEvent)

- GeoEvent instances input configuration use separate consumer groups:
  - With this configuration, all clients see all events
Stream Service resilience & scalability

10.6 best practice = multi-machine site of GeoEvent Servers

- Gateway is provided out-of-the-box at 10.6 for ingress:
  - all clients see all events by default
Stream Service transparency

10.6 best practice = multi-machine site of GeoEvent Servers

- A reverse proxy can be configured in between the clients and the stream services so that clients don’t have direct knowledge of the servers they are connecting to.
  - Example reverse proxies include NGiNX & Microsoft Application Request Routing (ARR).
5 Service Design Considerations
Service Design Considerations

Which would you choose?

Each “branch” in a service contains the same event data. In this example, with three branches, it is creating 3X the volume of data.

When possible, pre-filter the input data before ingesting.
Service Design Considerations
not all components are created equally

Which of these services will process the fastest?
Slowest?
Service Design Considerations

not all components are created equally

A

Input 1  ➔ (Service Area Calculator) ➔ Output 1

B

Input 1  ➔ (Field Enricher (Feature Service)) ➔ (Field Calculator) ➔ (Intersector) ➔ Output 1

C

Input 1  ➔ (Projector) ➔ (Simplifier) ➔ Output 1

The first service only contains components that are utilizing the internal service cache, which allows for the fastest processing.
Service Design Considerations

not all components are created equally

The second service modifies the incoming event geometry which can be “costly”.

These types of requests are typically very quick but can be impacted by geometry complexity.
Service Design Considerations

not all components are created equally

The third service utilizes Network Analyst to return a “drive time” polygon which can significantly impact throughput.
Service Design Considerations

other recommendations

• Configure Filters and/or Field Reducer Processors as early as possible in a service
  - This reduces the volume / data size of the events being processed
  - Potentially simplifies service configuration “down stream”

• Avoid Managed GeoEvent Definitions when possible
  - These are “system owned” definitions whose lifecycle is entirely controlled by the processors
  - Editing or Deleting a processor will remove these definitions
  - If necessary copy generated definition and edit processor to look for it

• Utilize the combination of Imported Definitions and Field Mapper Processor for Feature Service Outputs
  - This ensures that all of the event data is being written in the correct format
  - Can also be used to update only a portion of the fields
Upgrade Planning
Upgrade Planning

**what should be considered**

- **In-place Upgrade vs Clean Installation**
  - When possible do a clean install
  - GeoEvent Server install and uninstalls very quickly
- **Export Configuration & Global Settings from within GeoEvent Manager**
  - Use time to remove any unused definitions or components
- **Backup any configuration files that were modified in “/etc” folder**
- **Copy contents of “/deploy” folder (custom components)**
- **Delete contents of old site configuration (e.g. C:\arcgisserver\local\zookeeper)**
- **Install new version and import configurations**
Troubleshooting
Troubleshooting inputs

“I can’t get my data to come in…”

• Check the definition
  - Most input problems are with misconfigured schema (field names, data types, group structure)
  - Try letting GeoEvent Server create definition for you…
    - …but you will likely need to edit the definition and edit the input to use the one you modified
Troubleshooting outputs

“I can’t write my data to…”

- **Check the definition**
  - Most output problems are with misconfigured schema (field names, data types, group structure)
  - If possible import the definition from the service you are trying to write to
    - If an Esri Feature Service remove the reserved field names (e.g. ObjectID / OID)
  - Use the GeoEvent Logger application to verify the expected output data
Troubleshooting

backup

“Everything was working yesterday but…” or “Someone accidently deleted…”

• Did you make a backup of your configuration?
  - With ArcGIS GeoEvent Server 10.5 or newer, we did for you…

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Backup Settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backup Options</td>
<td>Back-up type: Daily, Properties: Time: 00:00</td>
<td></td>
</tr>
<tr>
<td>Enable Automatic Backups</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>Max Backups</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Edit Folder

Register a folder.

Name: Automatic Backups
Path: C:\ProgramData\Esri\GeoEvent
Help us improve the Real-Time & Big Data GIS Capabilities

http://esriurl.com/RealTimeSurvey
Questions / Feedback?

Adam Mollenkopf
Real-Time & Big Data GIS Capability Lead
amollenkopf@esri.com
@amollenkopf

Josh Joyner
ArcGIS GeoEvent Server Product Manager
jjoyner@esri.com