Real-Time & Big Data GIS:

Best Practices

Josh Joyner

Adam Mollenkopf

GIS INSPIRING WHAT'S NEXT

ArcGIS Enterprise

with real-time capabilities

Agenda:

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

1

Architecture Recommendations

GeoEvent Server

What are the primary factors I should consider?

- **Operating environment:** *m5.2xlarge*
  - **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.** *1 GBit/s*
- **Memory**
  - **size – 8GB has been required since 10.3.** *32GiB, default JVM max heap size is 4 GB*
  - **type – minimum of DDR3 is recommended.**
  - **clock speed (MHz) and transfer rate (Mbps) – the faster the better.**
- **Processors**
  - **# of cores – the more the better.** *8 vCPU*
  - **speed (GHz) – the faster the better.**
- **Disk**
  - **700MB required for installation** *10GB recommended minimum (new for 10.6)*
    - 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

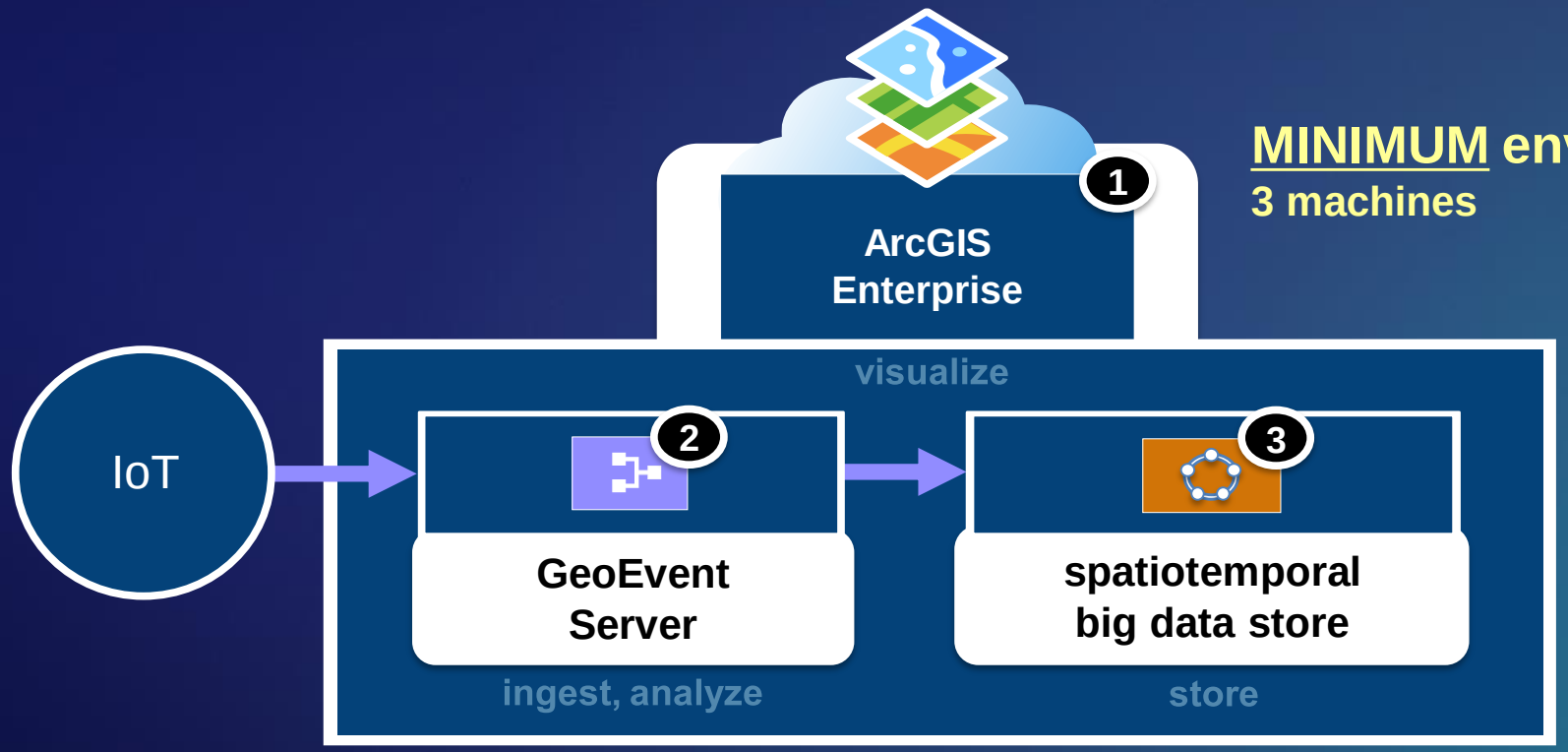
spatiotemporal big data store

What are the primary factors I should consider?

- **Operating environment:** *m5.2xlarge*
  - 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. *32GiB, big data store allocates 8GiB by default*
  - type – DDR3 is recommended.
  - clock speed (MHz) and transfer rate (Mbps) – the faster the better.
- **Processors**
  - # of cores – the more the better. *8 vCPU*
  - speed (GHz) – the faster the better.

ArcGIS Enterprise

with real-time capabilities



MINIMUM environment

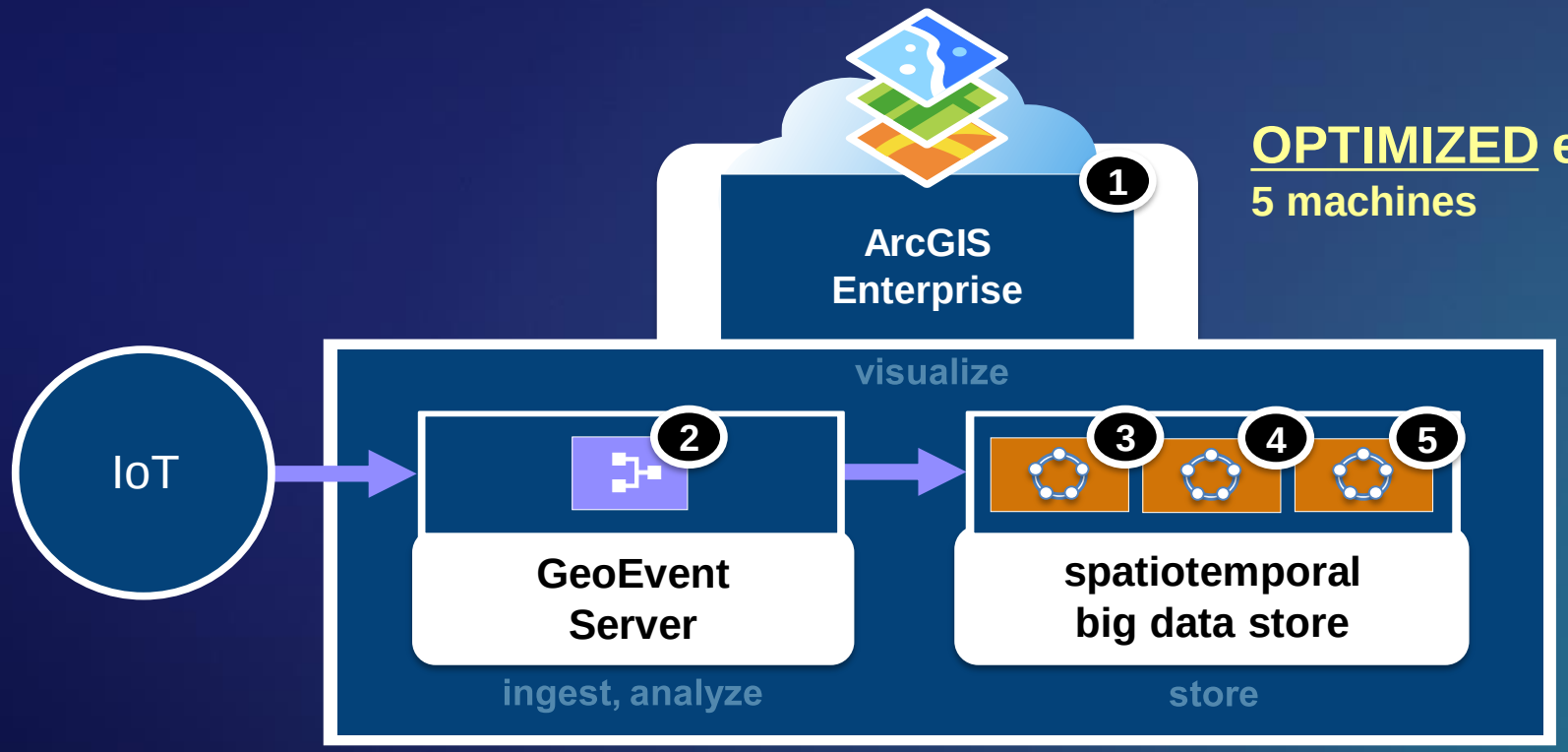
3 machines

functional servers & spatiotemporal big data store

SHOULD BE on ISOLATED machines

ArcGIS Enterprise

with real-time capabilities



OPTIMIZED environment

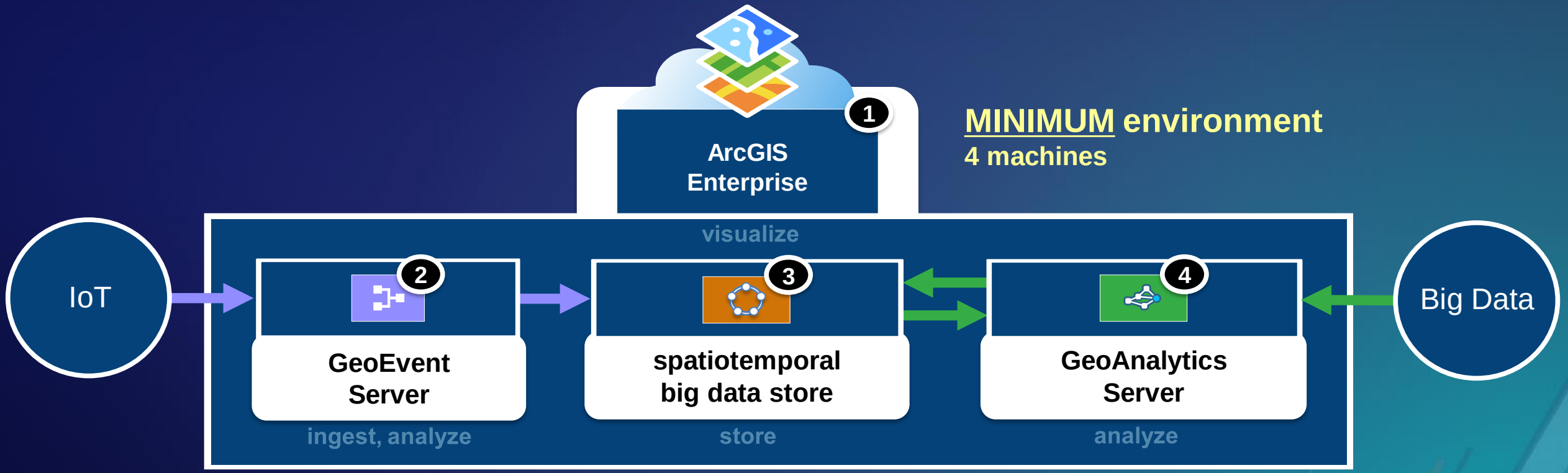
5 machines

functional servers & spatiotemporal big data store

SHOULD BE on ISOLATED machines

ArcGIS Enterprise

with real-time & big data GIS capabilities

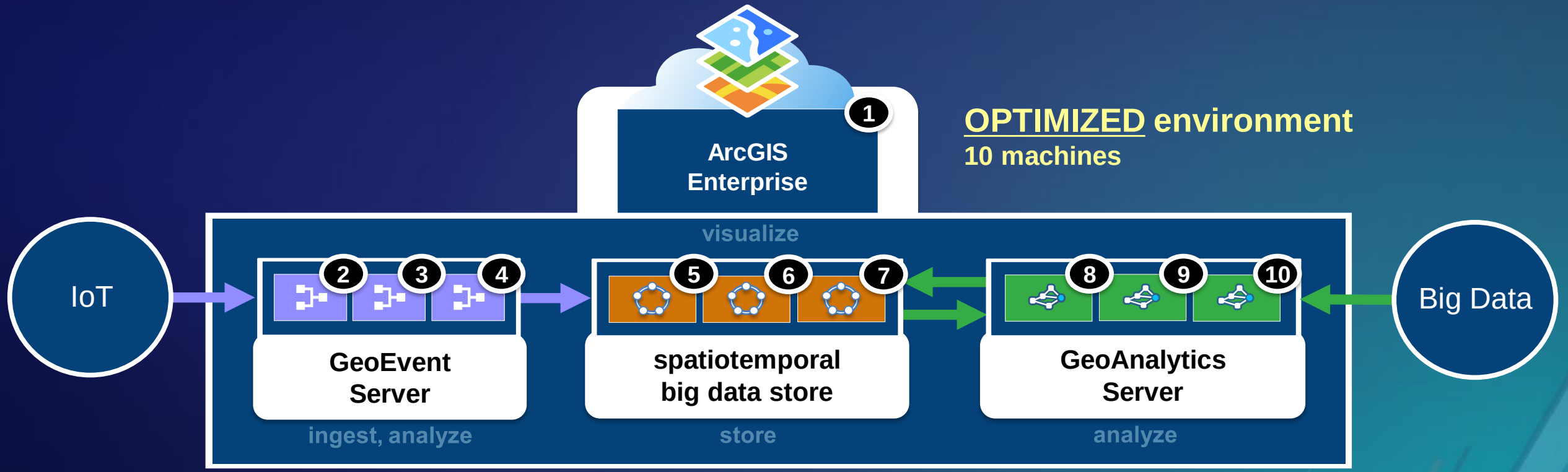


MINIMUM environment

4 machines

ArcGIS Enterprise

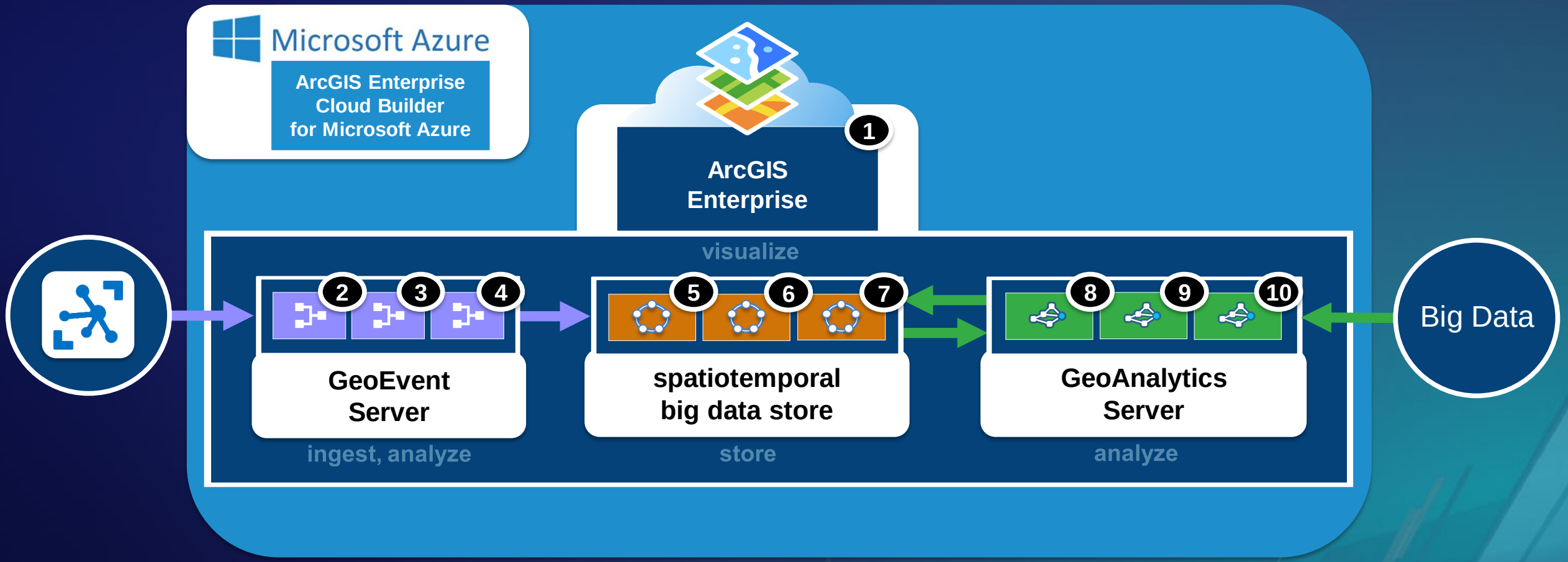
with real-time & big data GIS capabilities



ArcGIS Enterprise

with real-time & big data GIS capabilities on Microsoft Azure

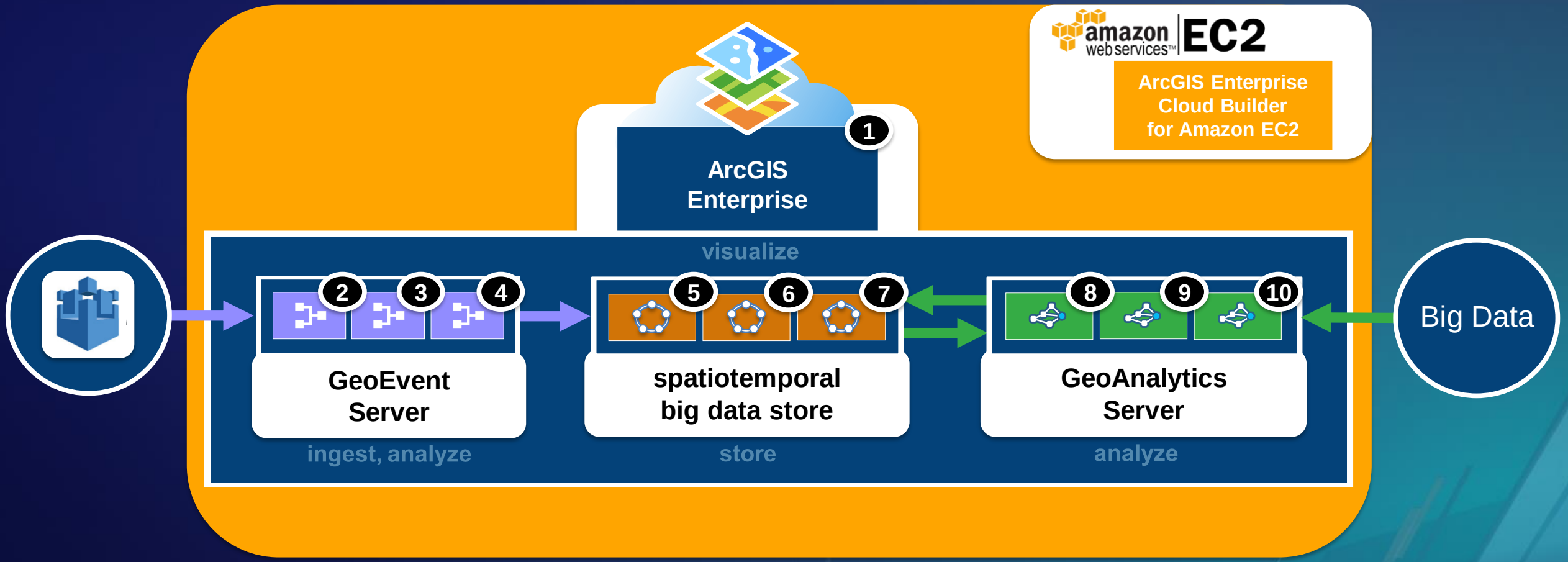
10.6.1



ArcGIS Enterprise

with real-time & big data GIS capabilities on Amazon EC2

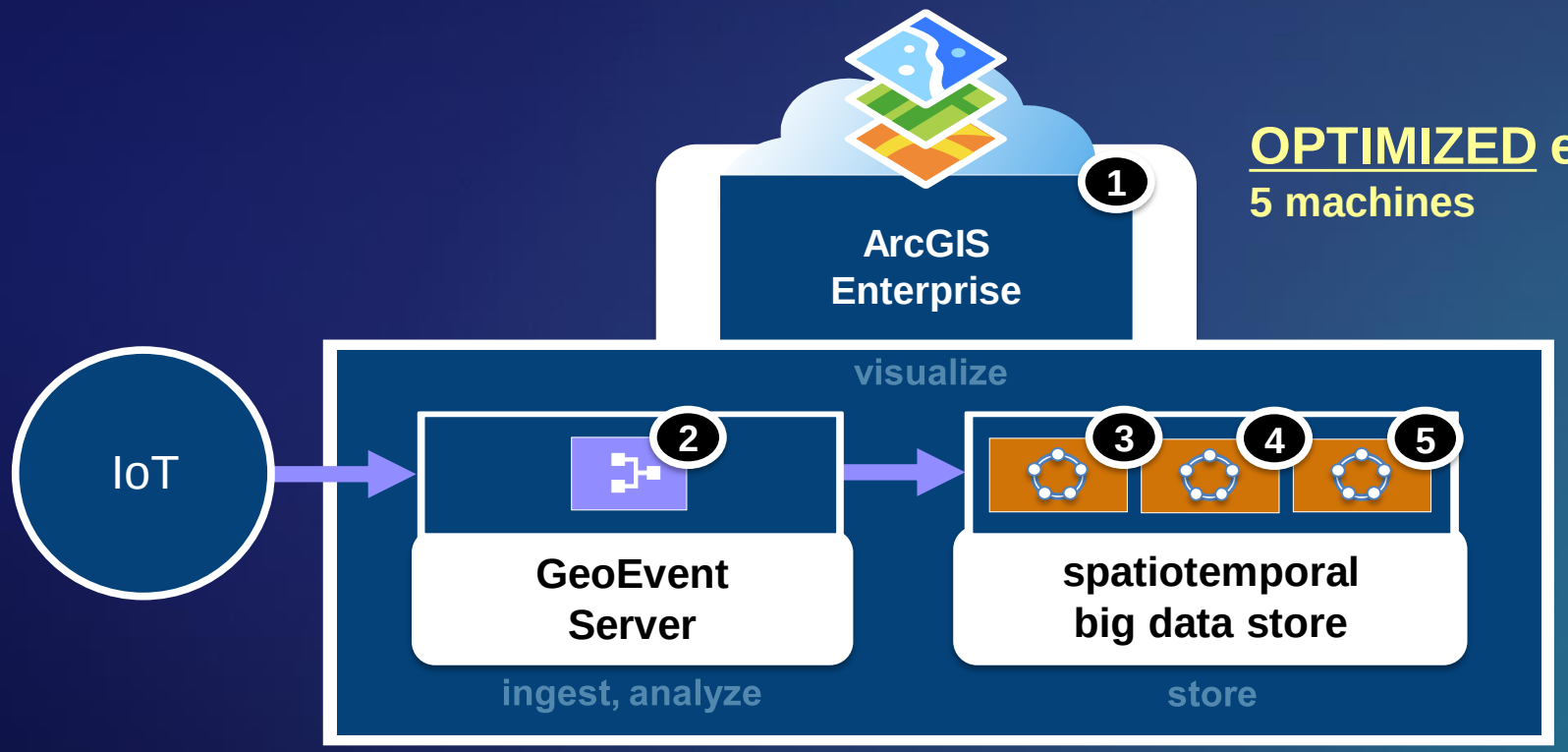
10.6.1



2 Big Data Storage

ArcGIS Enterprise

with real-time capabilities

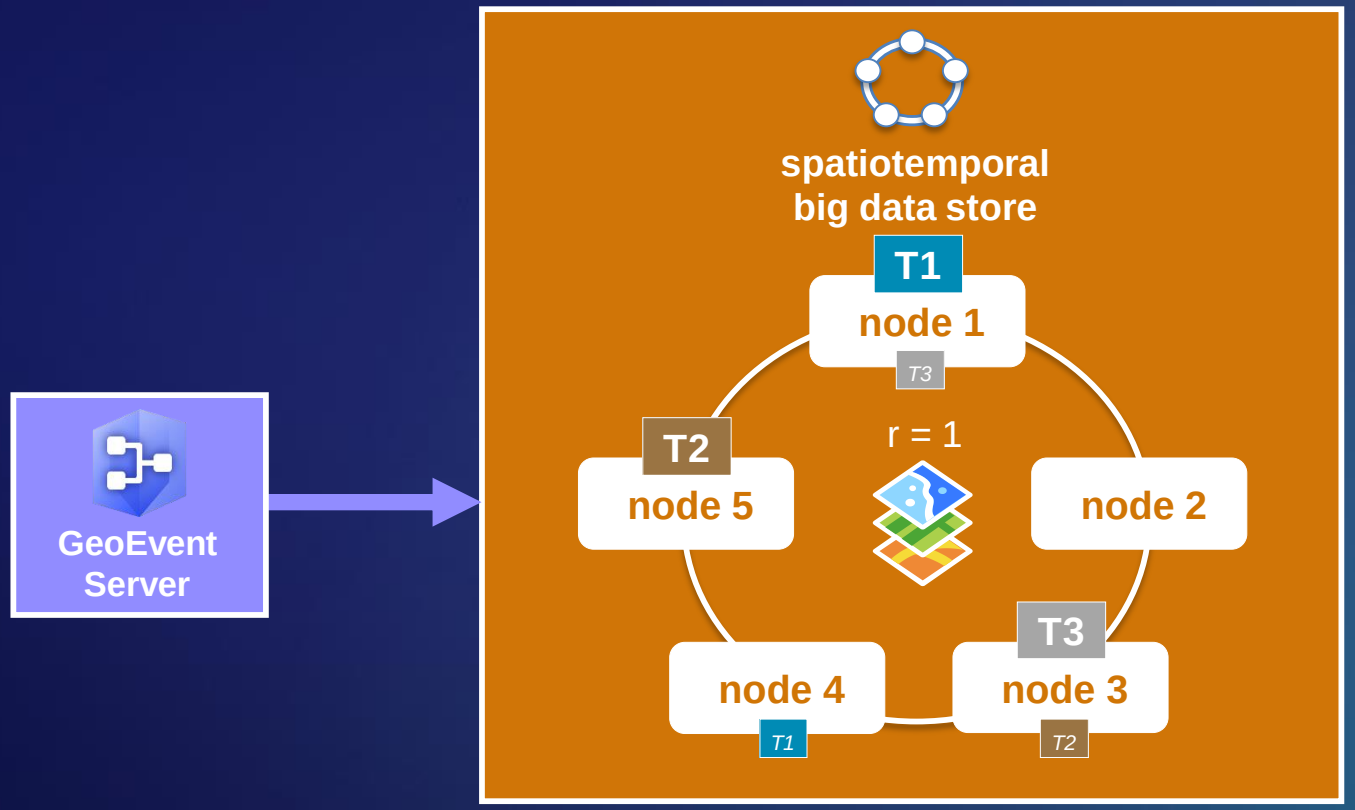


OPTIMIZED environment

5 machines

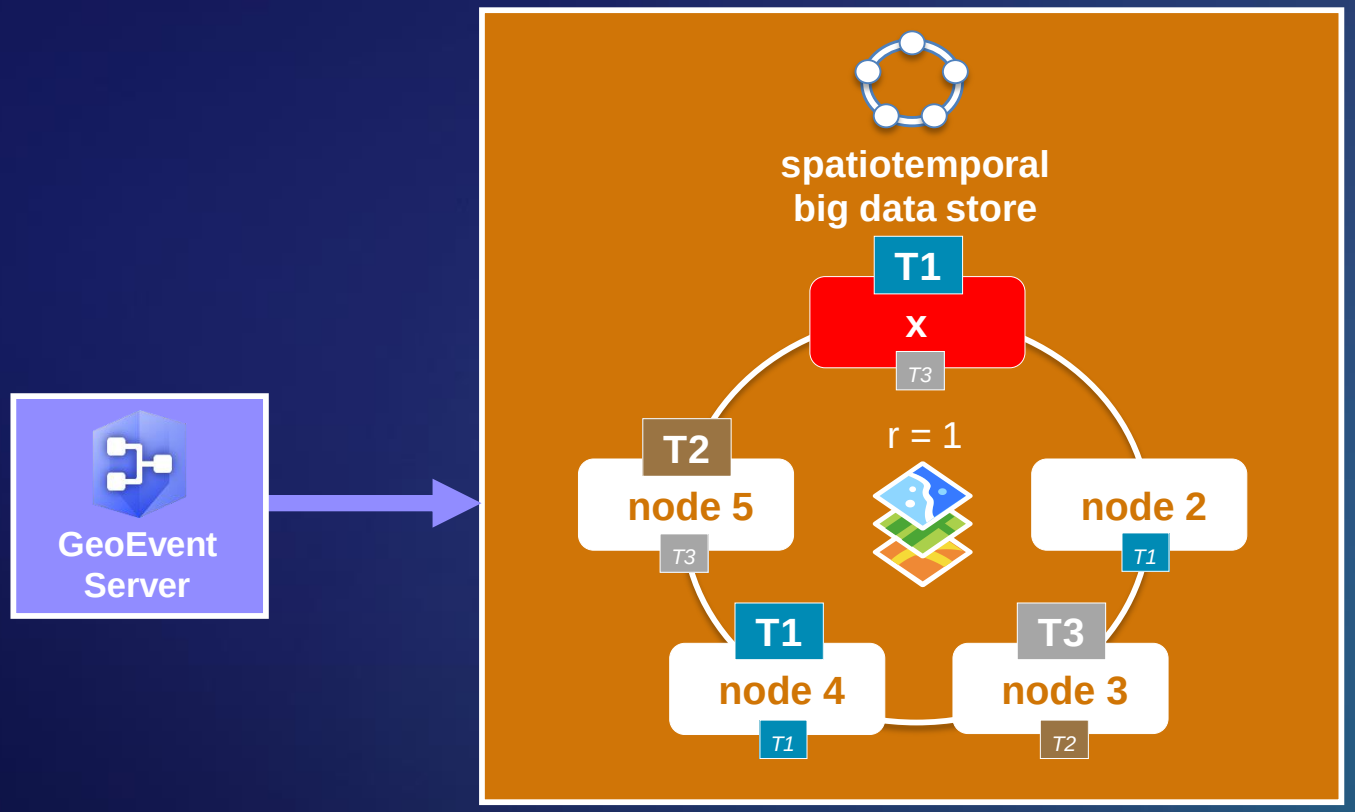
spatiotemporal big data store

shards & replication factor



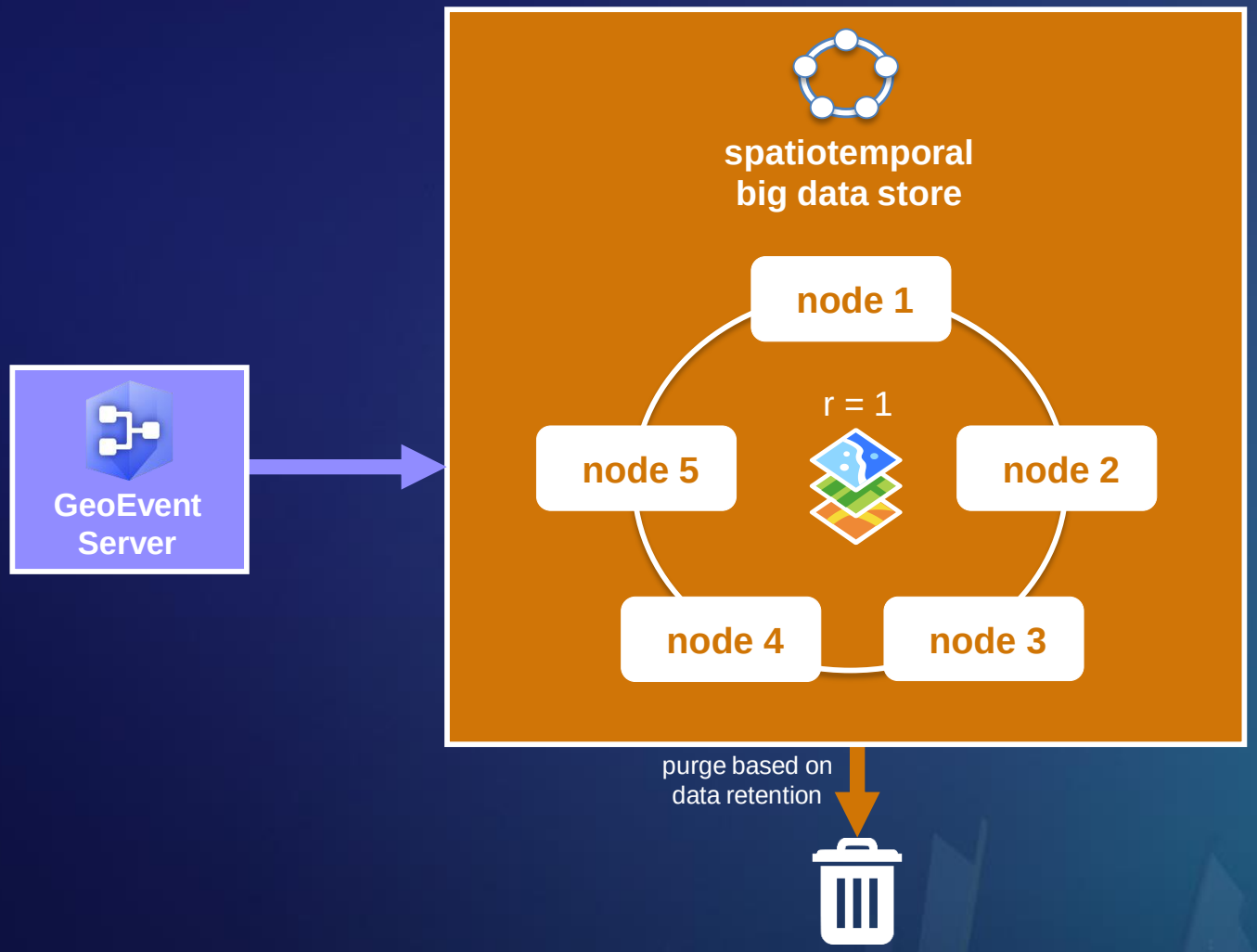
spatiotemporal big data store

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



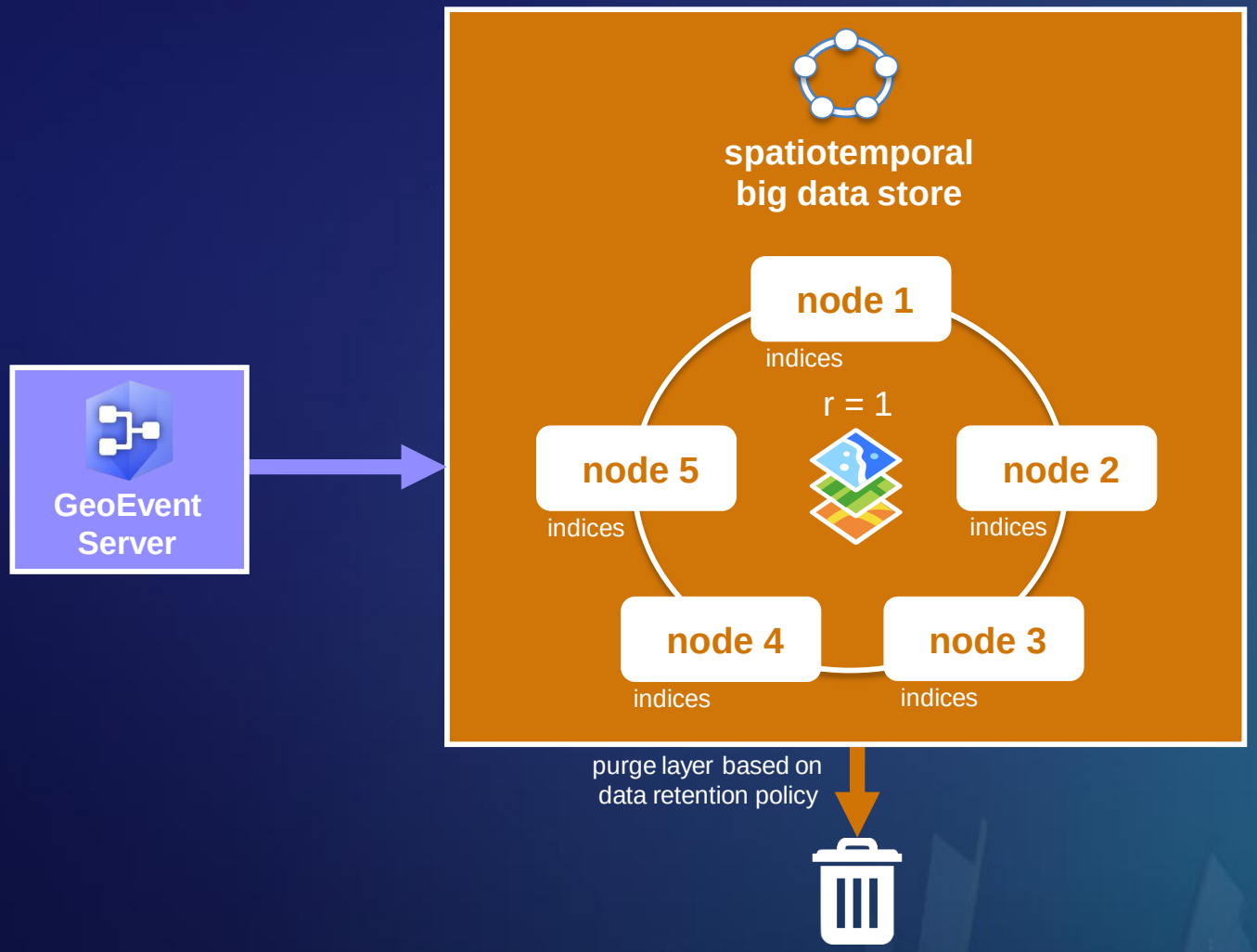
spatiotemporal big data store

data retention policies, configured per data source



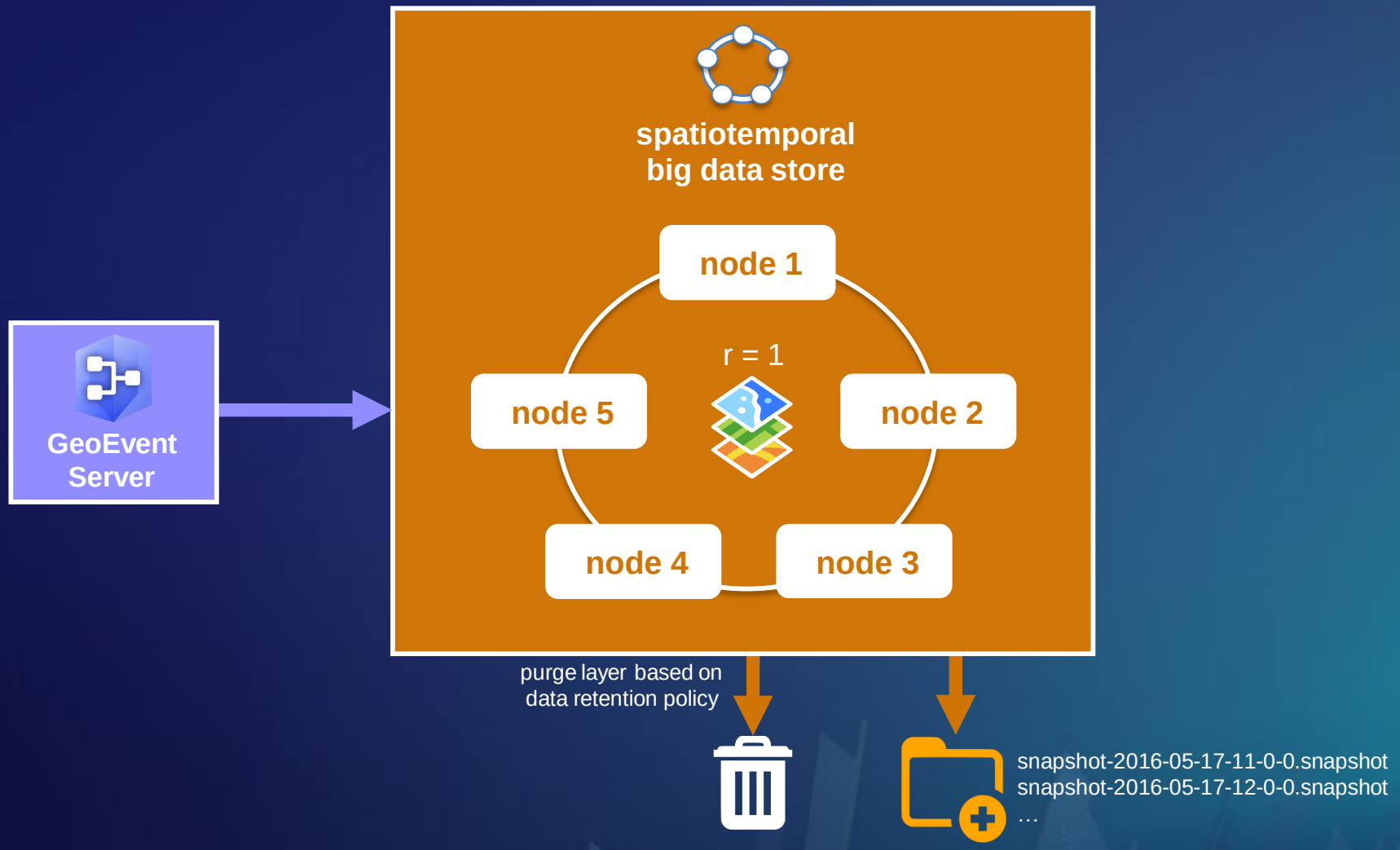
spatiotemporal big data store

rolling index option, set appropriately to the velocity of your observation data



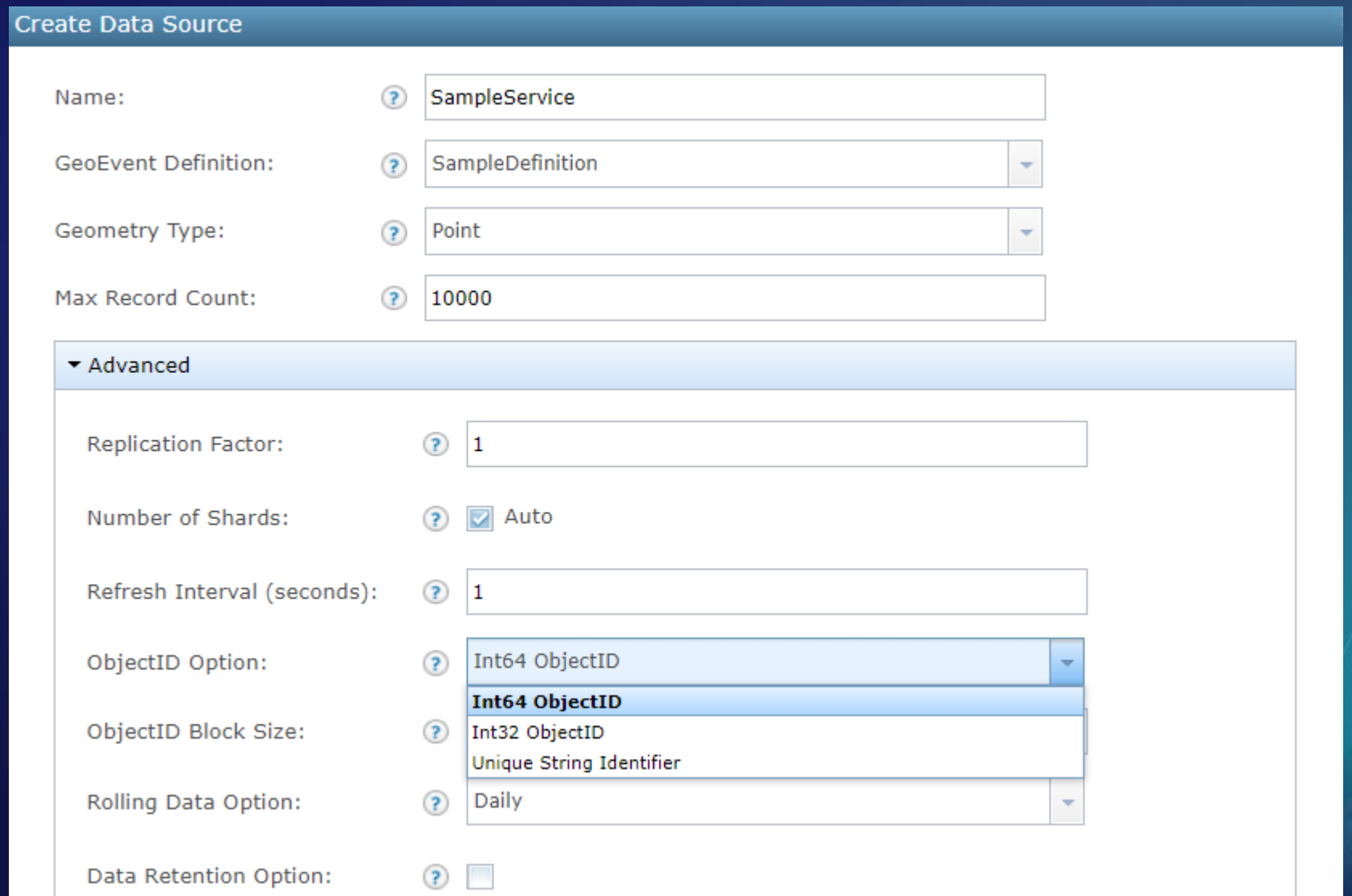
spatiotemporal big data store

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



spatiotemporal big data store

choosing an Object Id option



spatiotemporal big data store

choosing an Object Id option

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

IoT rate *per second*	per day	per week	per month	per year
1	86,400	604,800	2,592,000	31,104,000
10	864,000	6,048,000	25,920,000	311,040,000
100	8,640,000	60,480,000	259,200,000	3,110,400,000
1,000	86,400,000	604,800,000	2,592,000,000	31,104,000,000
10,000	864,000,000	6,048,000,000	25,920,000,000	311,040,000,000
100,000	8,640,000,000	60,480,000,000	259,200,000,000	3,110,400,000,000
1,000,000	86,400,000,000	604,800,000,000	2,592,000,000,000	31,104,000,000,000

spatiotemporal big data store

choosing an Object Id option

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

IoT rate *per second*	per day	per week	per month	per year	
1	86,400	604,800	2,592,000	31,104,000	
10	864,000	6,048,000	25,920,000	311,040,000	
100	8,640,000	60,480,000	259,200,000	3,110,400,000	
1,000	86,400,000	604,800,000	2,592,000,000	31,104,000,000	25 days
10,000	864,000,000	6,048,000,000	25,920,000,000	311,040,000,000	2.5 days
100,000	8,640,000,000	60,480,000,000	259,200,000,000	3,110,400,000,000	6 hours
1,000,000	86,400,000,000	604,800,000,000	2,592,000,000,000	31,104,000,000,000	36 minutes

spatiotemporal big data store

choosing an Object Id option

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

IoT rate *per second*	per day	per week	per month	per year
1	86,400	604,800	2,592,000	31,104,000
10	864,000	6,048,000	25,920,000	311,040,000
100	8,640,000	60,480,000	259,200,000	3,110,400,000
1,000	86,400,000	604,800,000	2,592,000,000	31,104,000,000
10,000	864,000,000	6,048,000,000	25,920,000,000	311,040,000,000
100,000	8,640,000,000	60,480,000,000	259,200,000,000	3,110,400,000,000
1,000,000	86,400,000,000	604,800,000,000	2,592,000,000,000	31,104,000,000,000

spatiotemporal big data store

choosing an Object Id option

	Max Value	# of IDs	ArcGIS Clients
Int32	2,147,483,647	2.1 billion	Pro, Desktop, Ops Dashboard, …
Int64 (signed)	9,223,372,036,854,775,807	9.2 quintillion	JavaScript, custom apps
UniqueStringID	n/a	unlimited	JavaScript, custom apps

IoT rate *per second*	per day	per week	per month	per year
1	86,400	604,800	2,592,000	31,104,000
10	864,000	6,048,000	25,920,000	311,040,000
100	8,640,000	60,480,000	259,200,000	3,110,400,000
1,000	86,400,000	604,800,000	2,592,000,000	31,104,000,000
10,000	864,000,000	6,048,000,000	25,920,000,000	311,040,000,000
100,000	8,640,000,000	60,480,000,000	259,200,000,000	3,110,400,000,000
1,000,000	86,400,000,000	604,800,000,000	2,592,000,000,000	31,104,000,000,000

3 Performance, Resiliency, & Scalability

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)**

ArcGIS GeoEvent Server	10.2	10.3	10.4	10.5	10.6
Velocity throughput *measured in events per second (e/s)*	*up to* 500 *e/s*	*up to* 2,000 *e/s*	*up to* 3,000 *e/s*	*up to* 4,000 *e/s*	*up to* 6,000 *e/s*

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)**

ArcGIS GeoEvent Server	10.2	10.3	10.4	10.5	10.6	10.6.1
Velocity throughput *measured in events per second (e/s)*	*up to* 500 *e/s*	*up to* 2,000 *e/s*	*up to* 3,000 *e/s*	*up to* 4,000 *e/s*	*up to* 6,000 *e/s*	*up to* 10,000 *e/s*

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

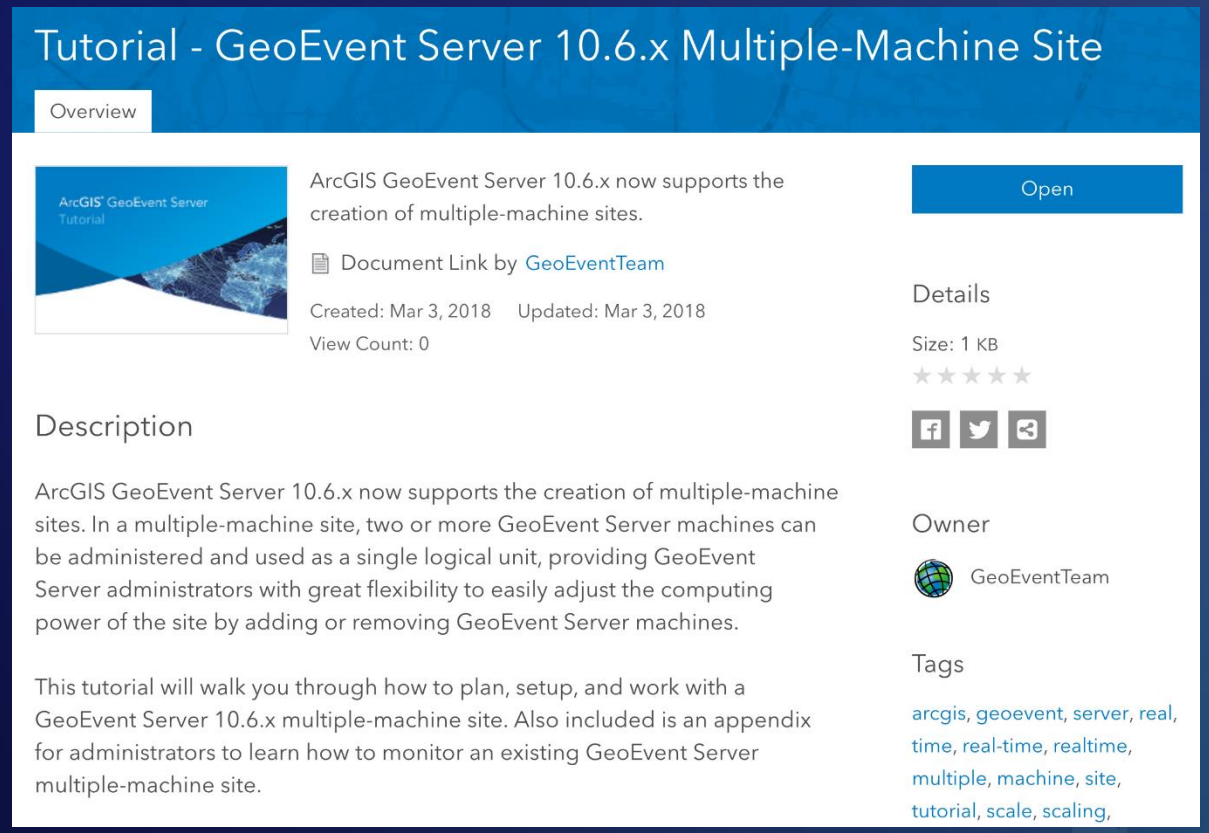
- Provides users with a resilient & scalable Real-Time GIS deployment OUT-OF-THE-BOX.
  - Introduces a gateway process that is automatically configured as part of GeoEvent Server installation.
- Provides flexibility for all input types.

ArcGIS GeoEvent Server	10.2	10.3	10.4	10.5	10.6	10.6.1
Velocity throughput *measured in events per second (e/s)*	*up to* 500 *e/s*	*up to* 2,000 *e/s*	*up to* 3,000 *e/s*	*up to* 4,000 *e/s*	*up to* 6,000 *e/s*	*up to* 10,000 *e/s*
Resiliency & Scalability *via multi-machine site*	no	no	no	no	yes	yes

Performance, Resiliency & Scalability

multi-machine site support

- Available Now: http://links.esri.com/geoevent-multiplemachine



Tutorial - GeoEvent Server 10.6.x Multiple-Machine Site

Overview

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

Document Link by GeoEventTeam

Created: Mar 3, 2018 Updated: Mar 3, 2018

View Count: 0

Open

Details

Size: 1 KB

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.

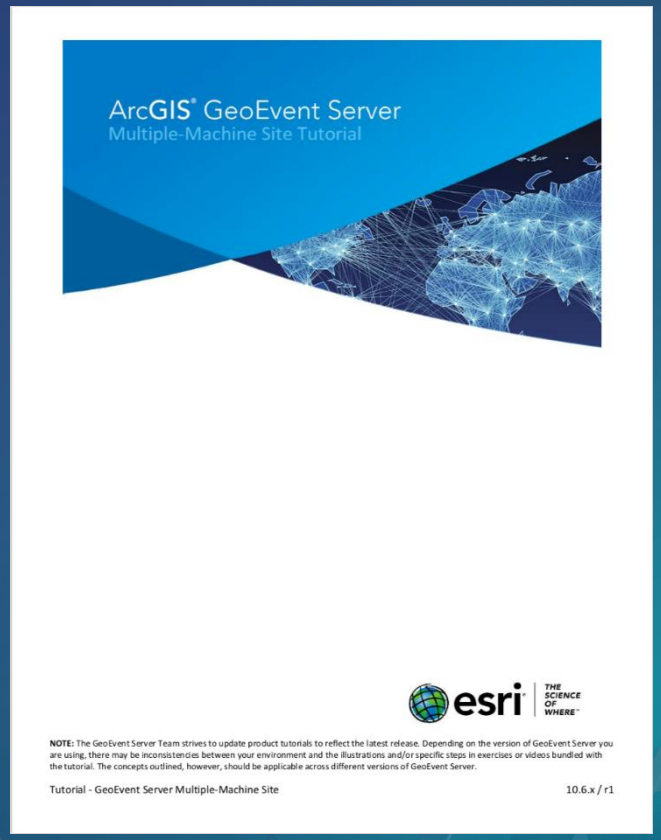
This tutorial will walk you through how to plan, setup, and work with a GeoEvent Server 10.6.x multiple-machine site. Also included is an appendix for administrators to learn how to monitor an existing GeoEvent Server multiple-machine site.

Owner

GeoEventTeam

Tags

arcgis, geoevent, server, real, time, real-time, realtime, multiple, machine, site, tutorial, scale, scaling,



ArcGIS® GeoEvent Server

Multiple-Machine Site Tutorial

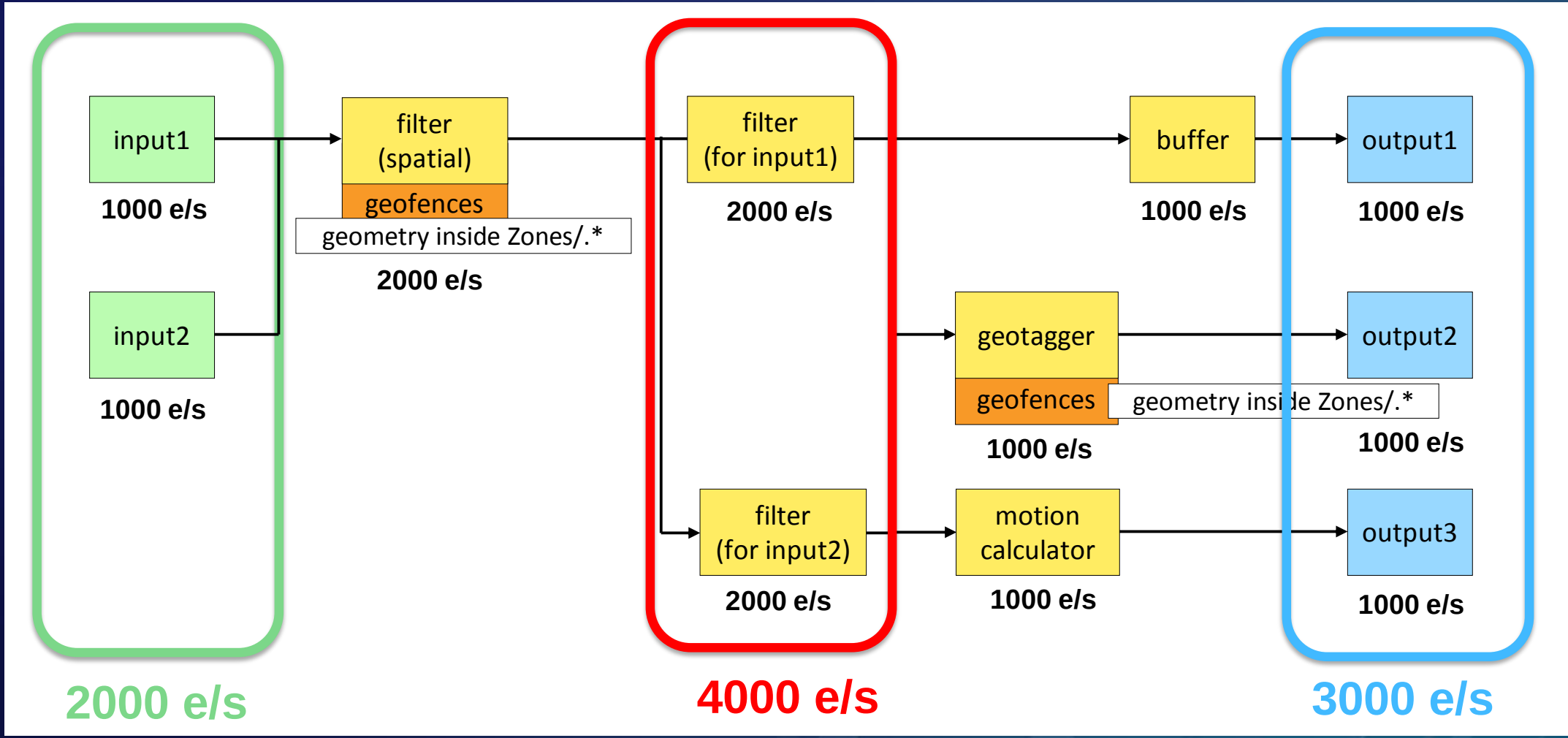
esri THE SCIENCE OF WHERE

NOTE: The GeoEvent Server Team strives to update product tutorials to reflect the latest release. Depending on the version of GeoEvent Server you are using, there may be inconsistencies between your environment and the illustrations and/or specific steps in exercises or videos bundled with the tutorial. The concepts outlined, however, should be applicable across different versions of GeoEvent Server.

Tutorial - GeoEvent Server Multiple-Machine Site 10.6.x / r1

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)

```
wrapper.java.additional.12=-XX:+CMSClassUnloadingEnabled

# Maximum Java Heap Size (in MB). Note: use this option verse the one below 'wrapper.java.maxmemory' when setting the max heap size.
wrapper.java.additional.13=-Xmx4096m

# Initial Java Heap Size (in MB)
#wrapper.java.initmemory=128
```

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"

```
com.esri.ges.manager.servicemanager.maxCacheSize=1000
```

- You may also need to modify the Incident Manager Setting in the Global Setting Tab if used in conjunction with the Incident Detector Processor

Incident Manager Settings		
Maximum number of closed incidents	1000	
Maximum number of opened incidents	1000	

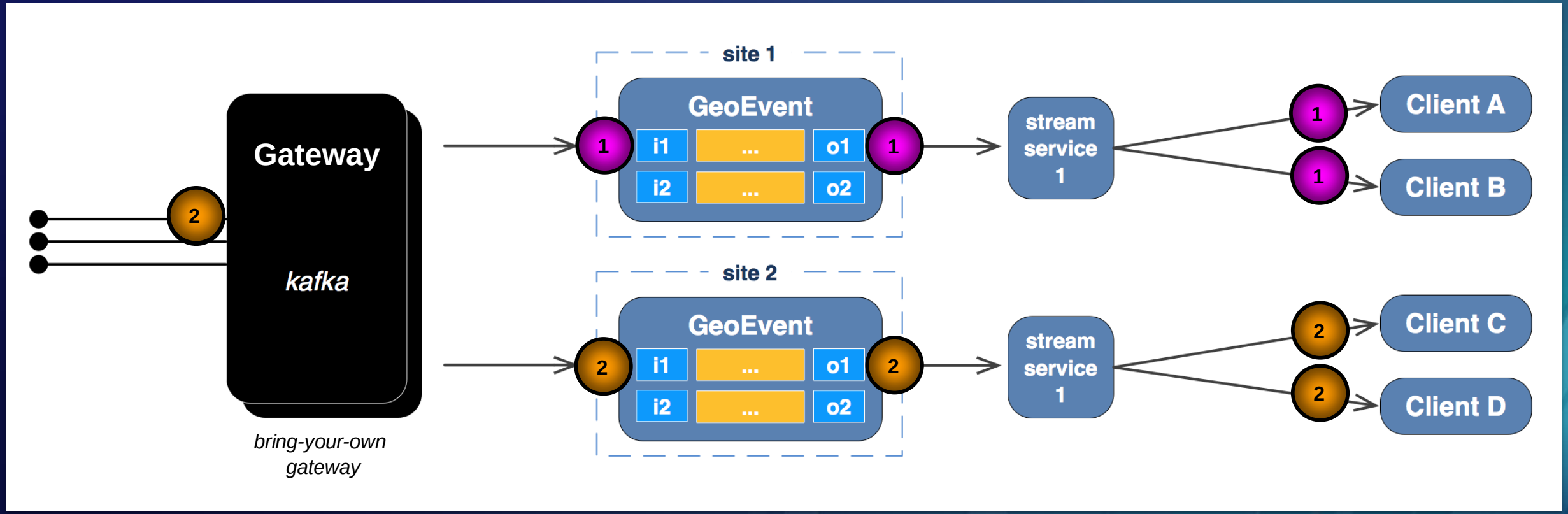
4 Stream Services

Stream Service resilience & scalability

10.5

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

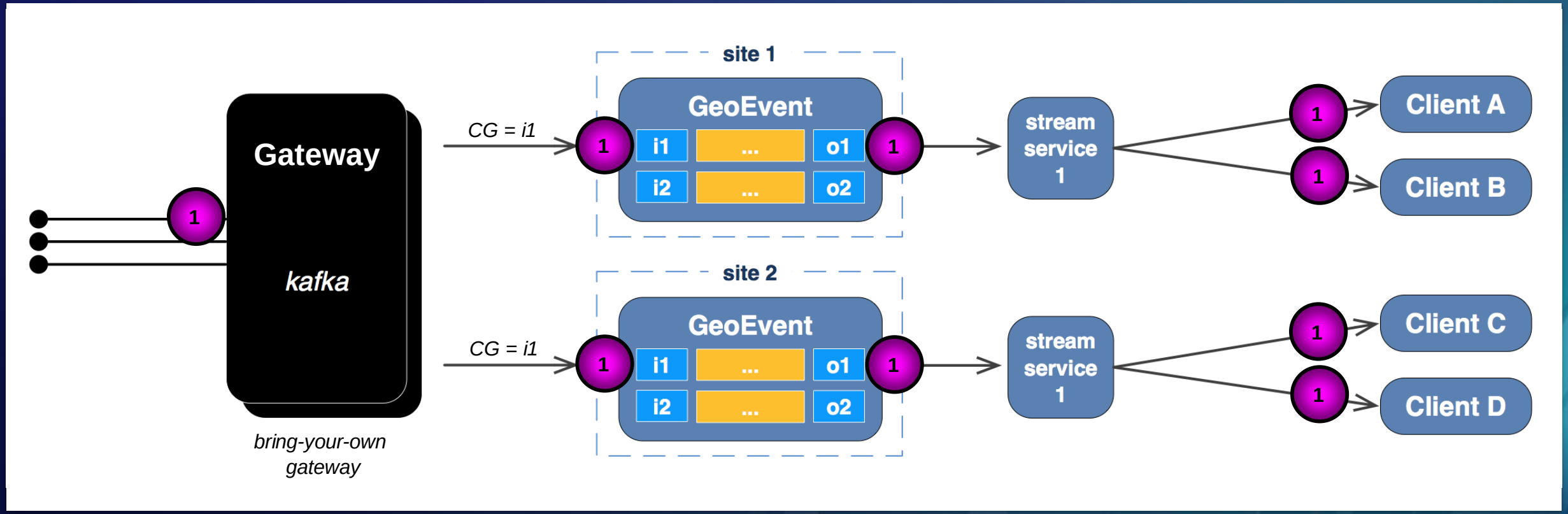


Stream Service resilience & scalability

10.5

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

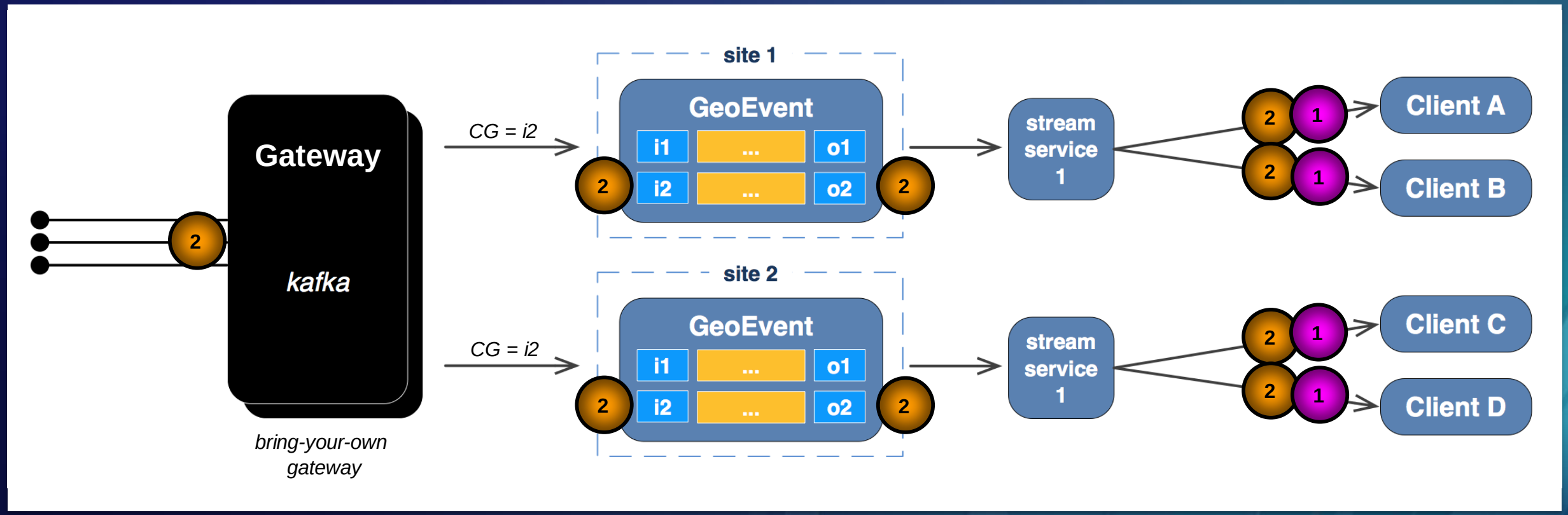


Stream Service resilience & scalability

10.5

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

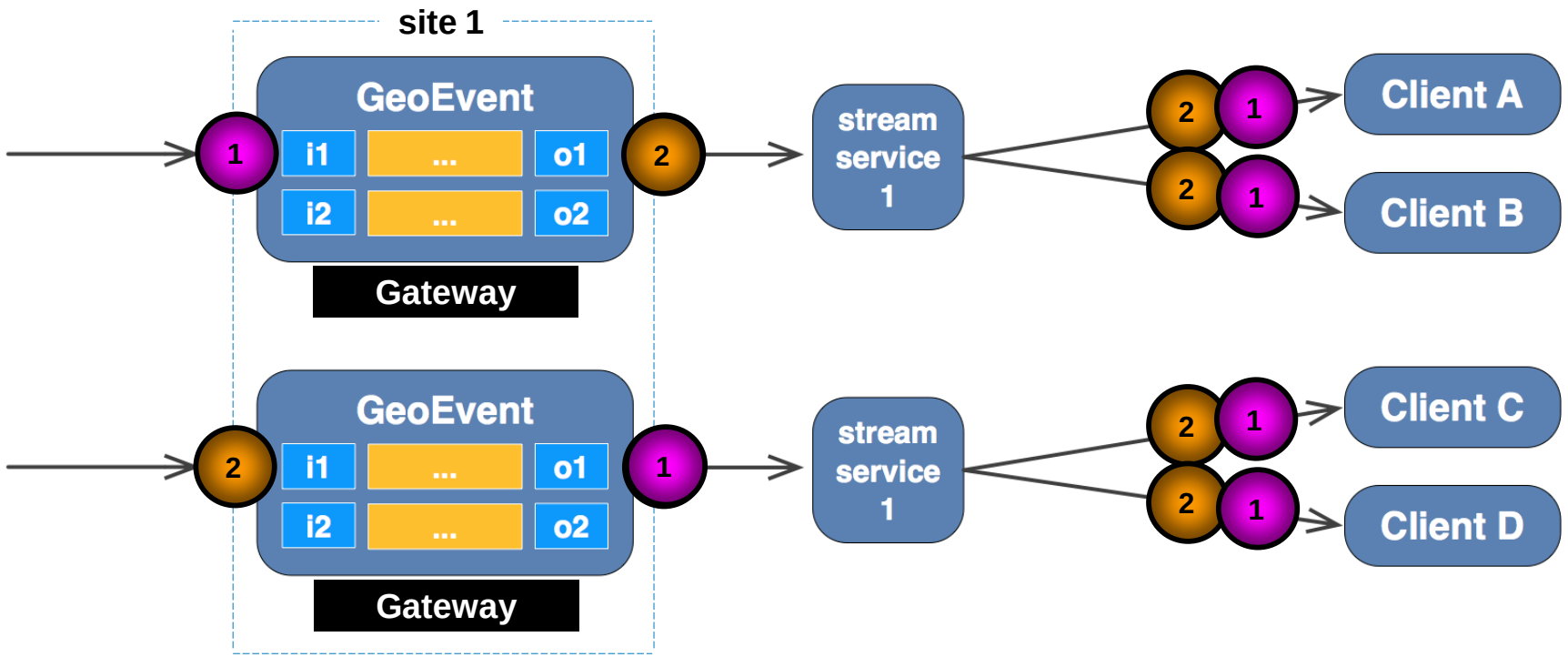


Stream Service resilience & scalability

10.6

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

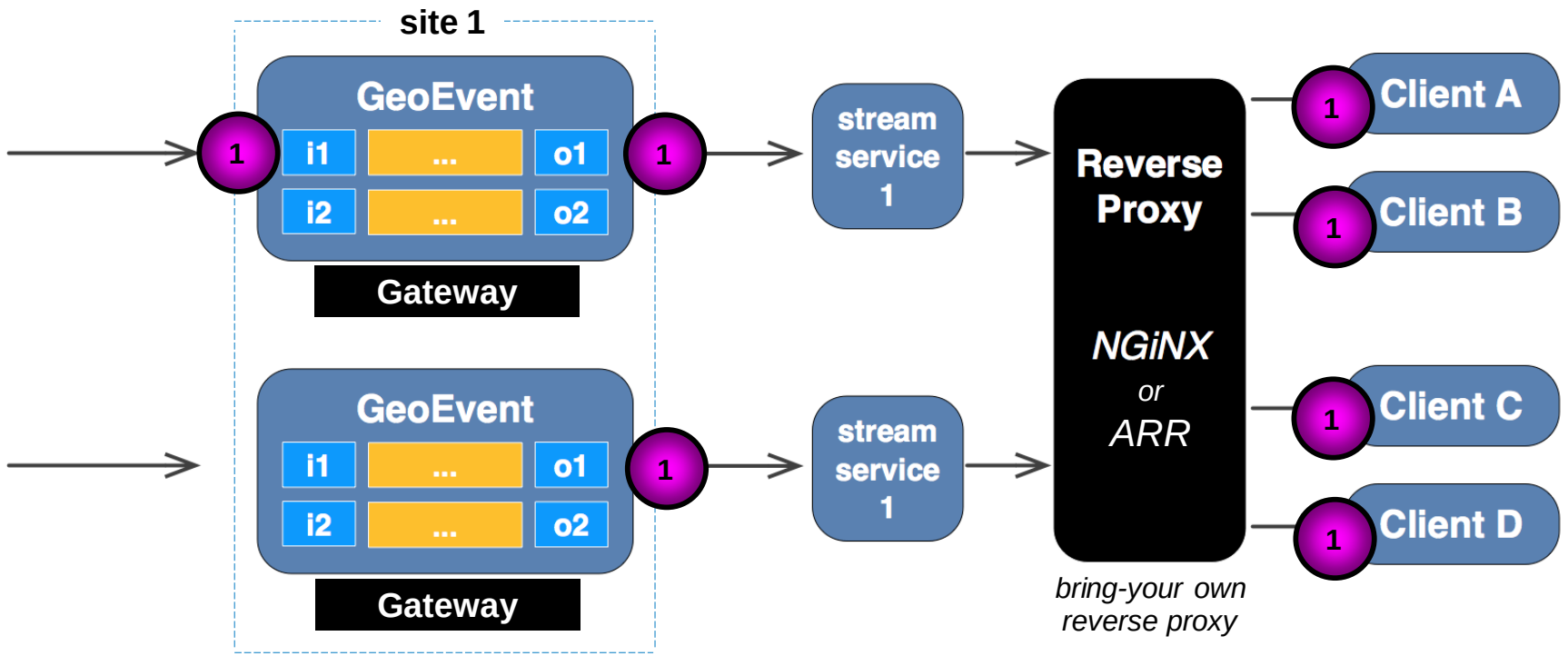


Stream Service transparency

10.6 best practice = multi-machine site of GeoEvent Servers

10.6

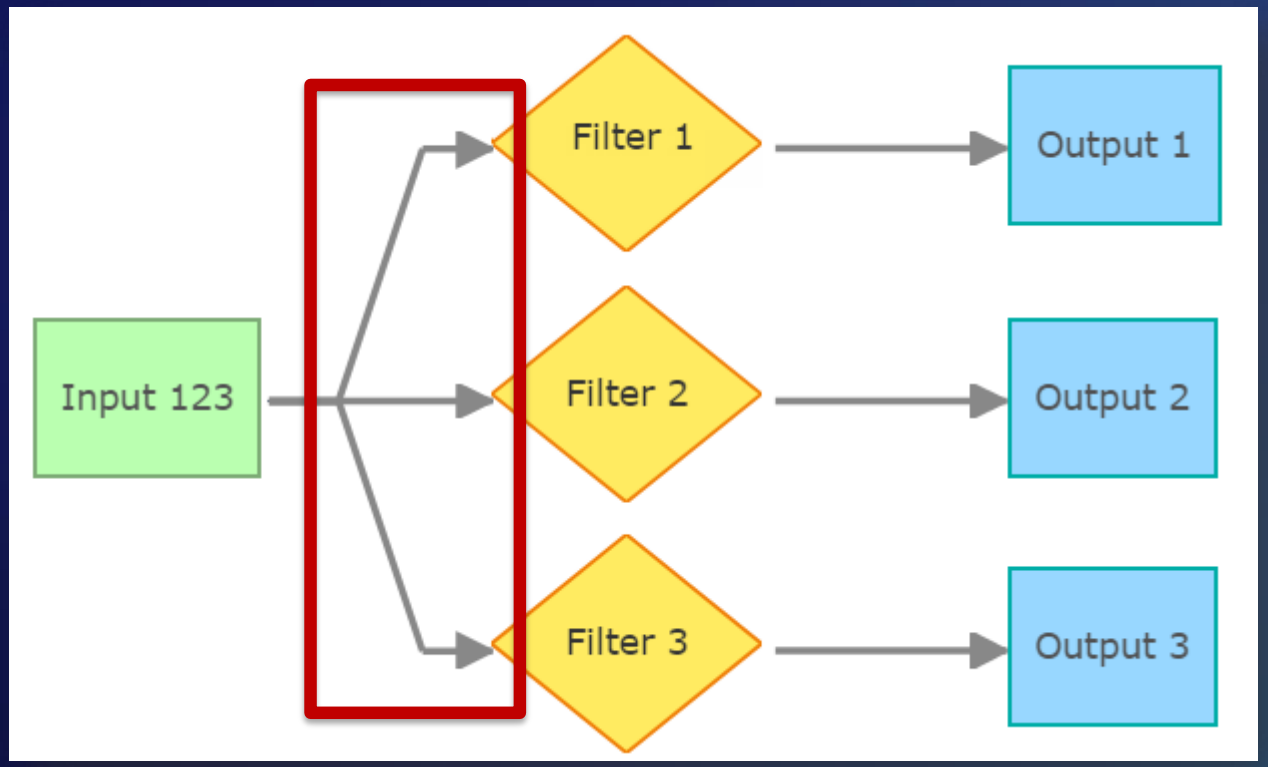
- A reverse proxy can be configured in between the clients and the stream services so that clients don't have direct knowledge of the servers they are connecting to.
  - Example reverse proxies include NGiNX & Microsoft Application Request Routing (ARR).



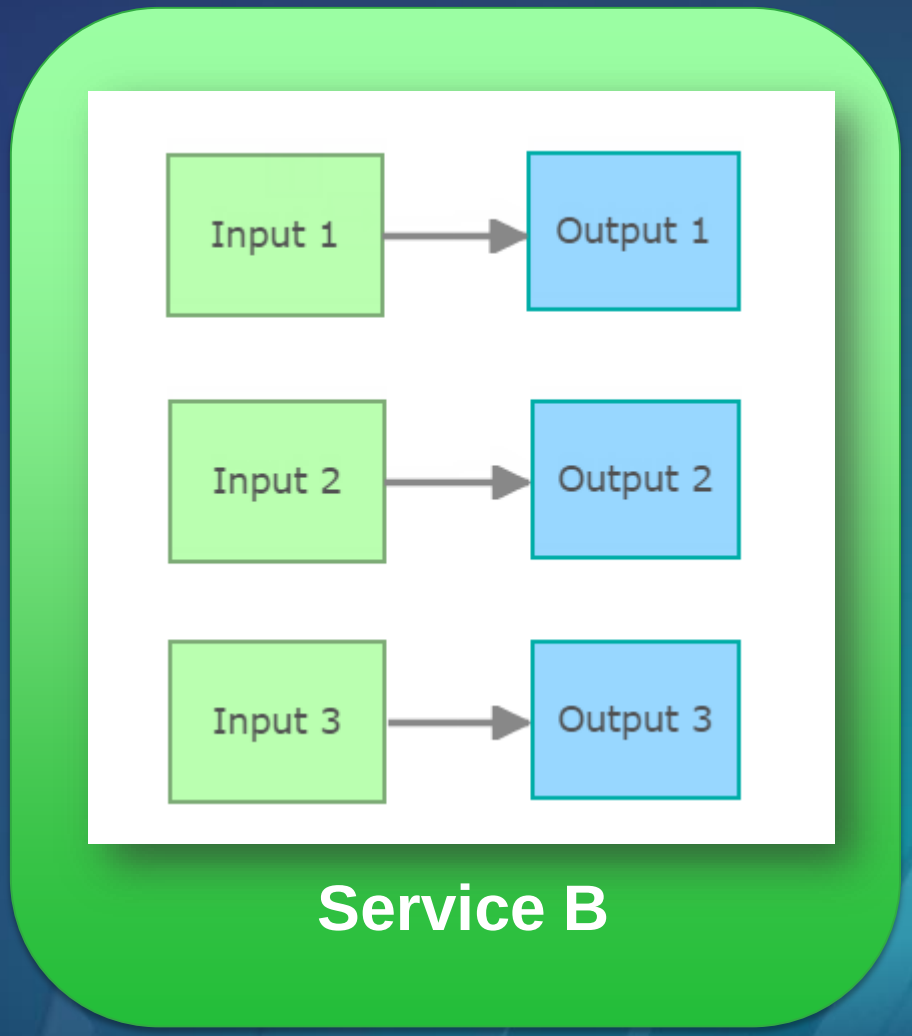
5 Service Design Considerations

Service Design Considerations

which would you choose?



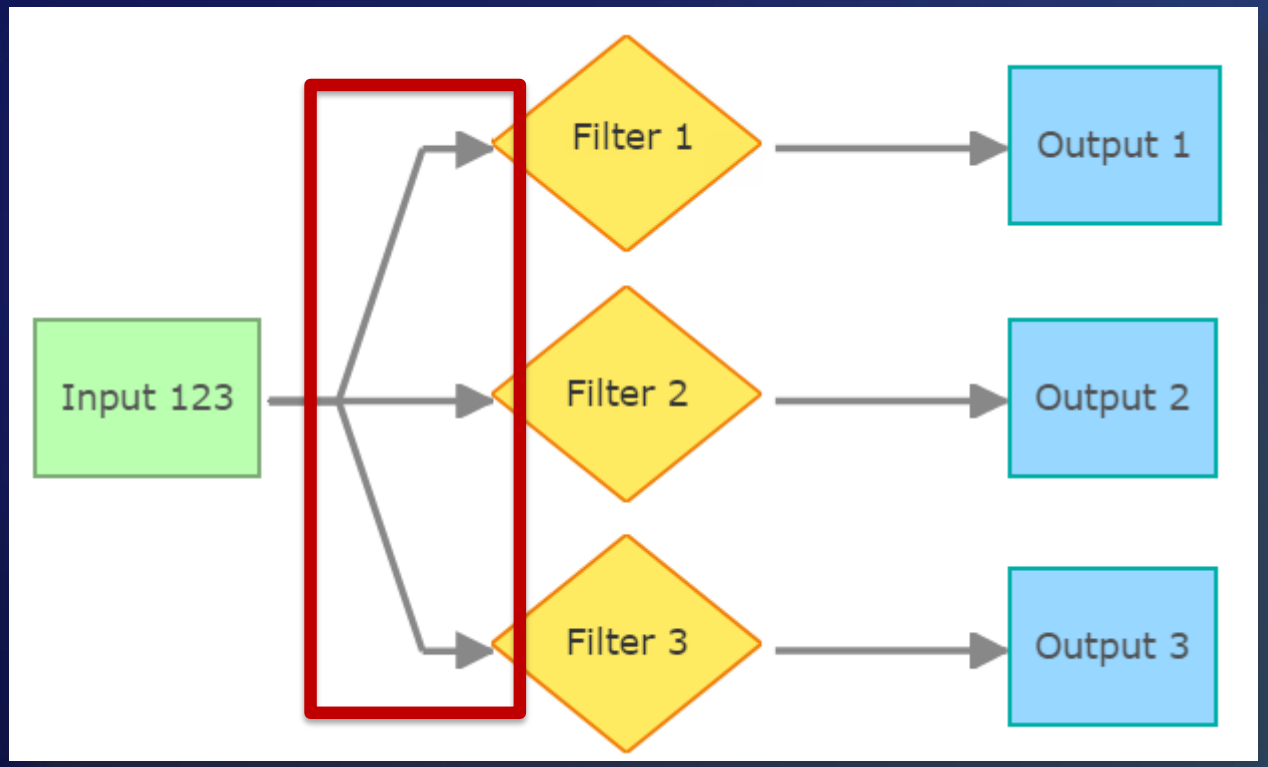
Service A



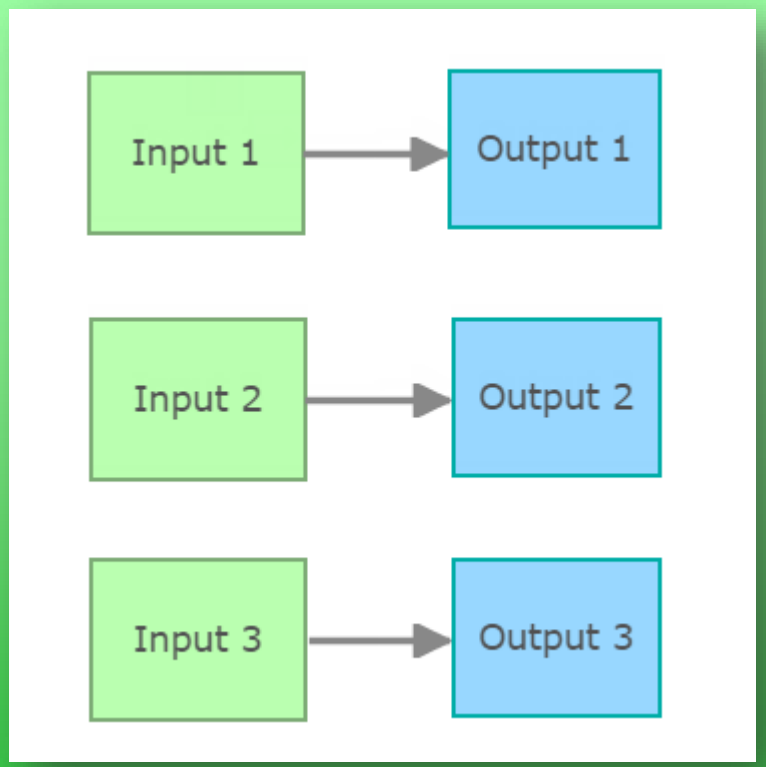
Service B

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.

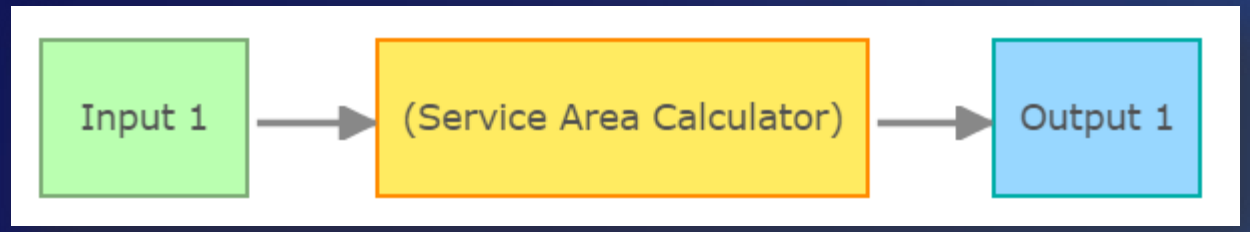


When possible, pre-filter the input data before ingesting.

Service Design Considerations

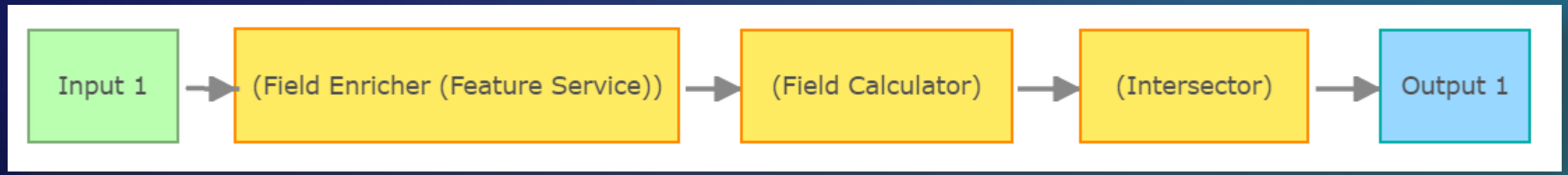
not all components are created equally

A

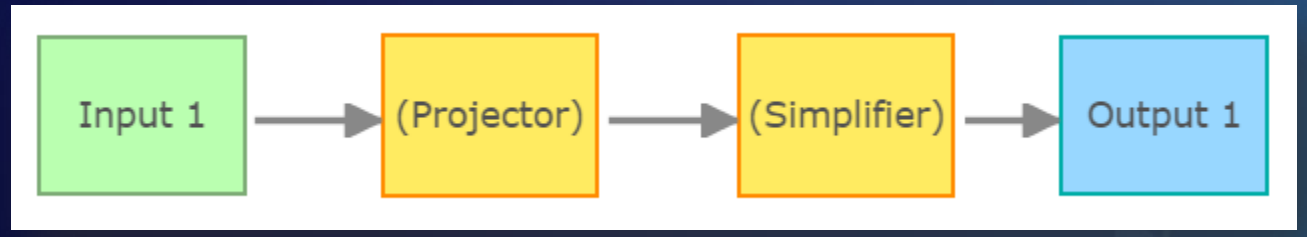


Which of these services will process the fastest? Slowest?

B



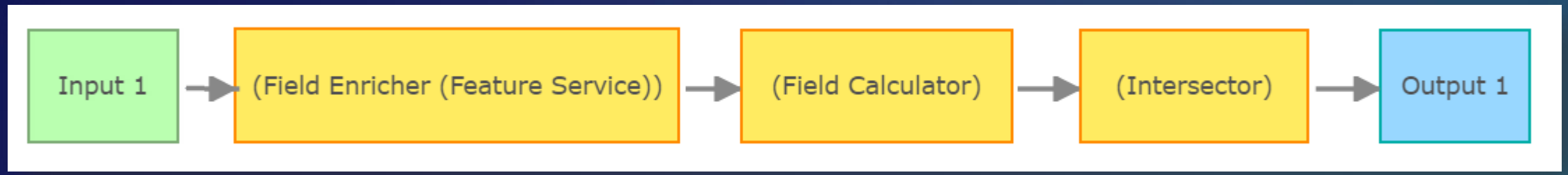
C



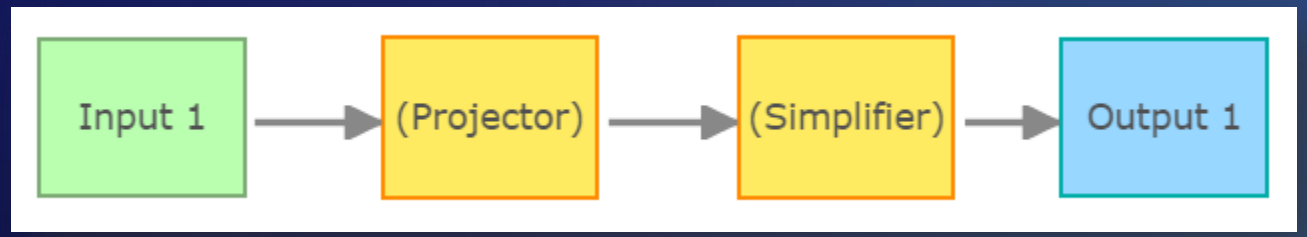
Service Design Considerations

not all components are created equally

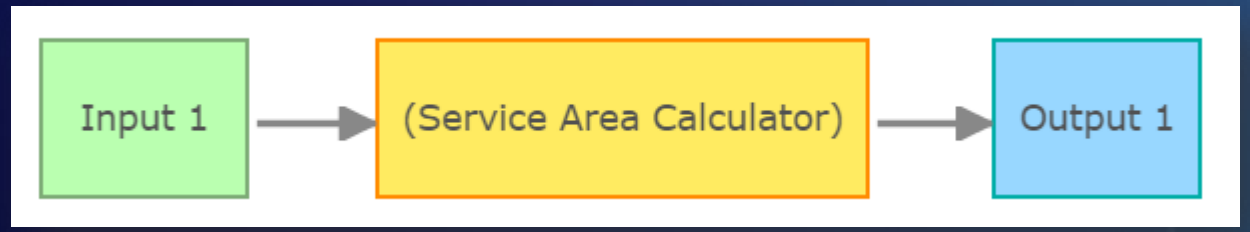
B



C



A

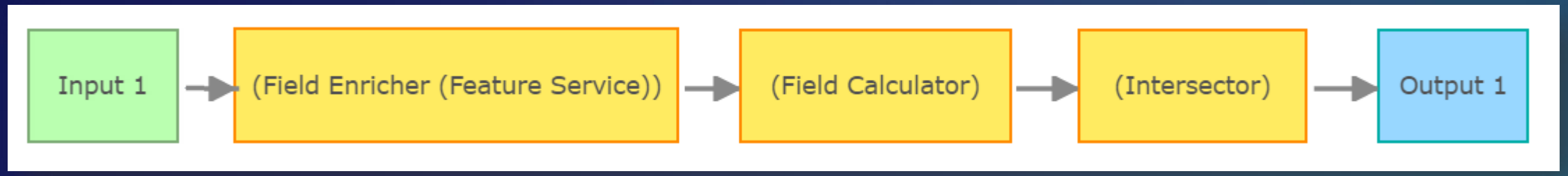


The first service only contains components that are utilizing the internal service cache, which allows for the fastest processing.

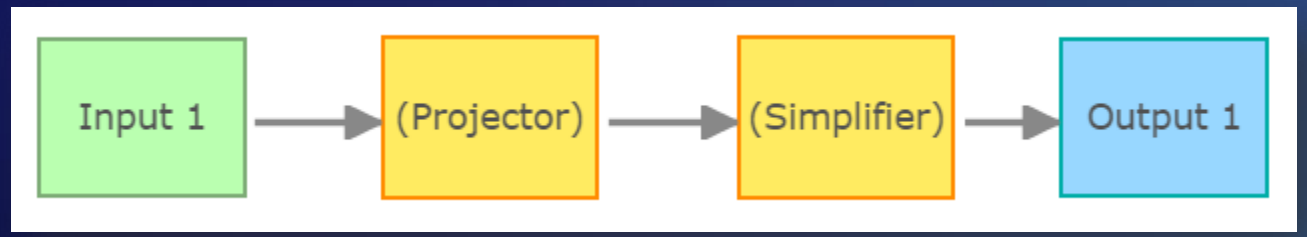
Service Design Considerations

not all components are created equally

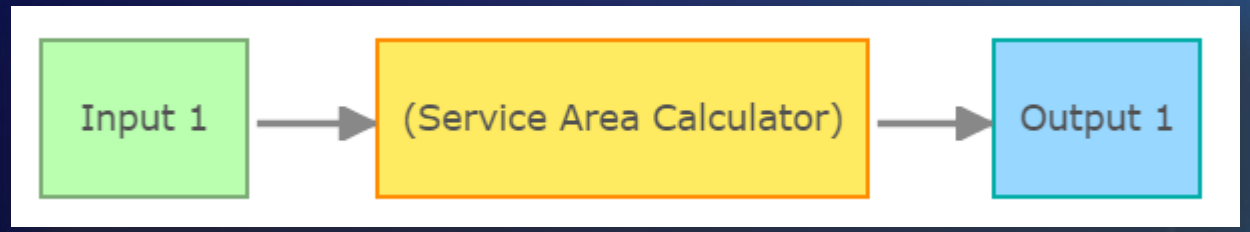
B



C



A



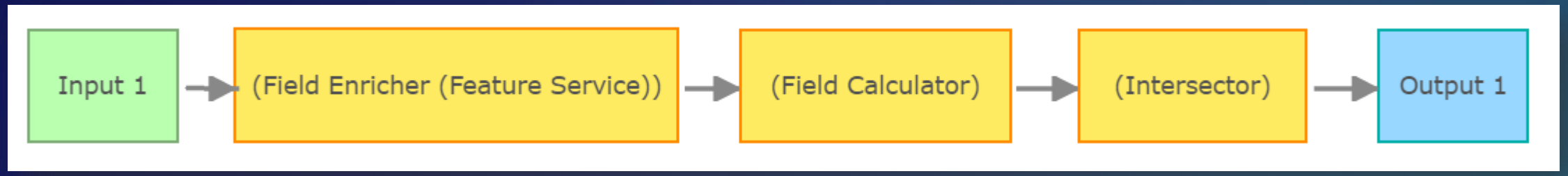
The second service modifies the incoming event geometry which can be “costly”.

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

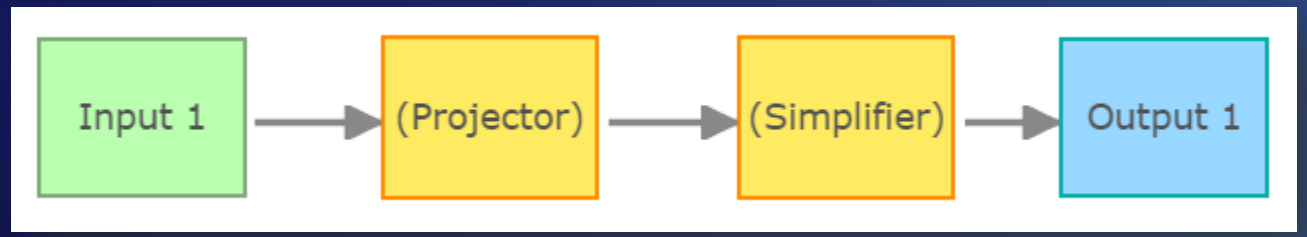
Service Design Considerations

not all components are created equally

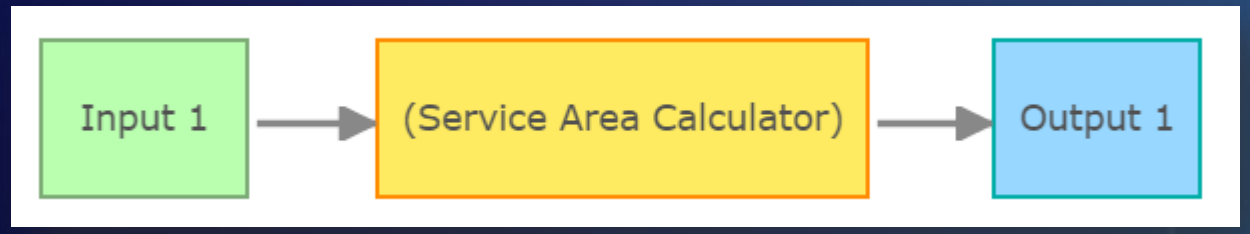
B



C



A



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**

6 Upgrade Planning

Upgrade Planning

what should be consider

- 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

7 Troubleshooting

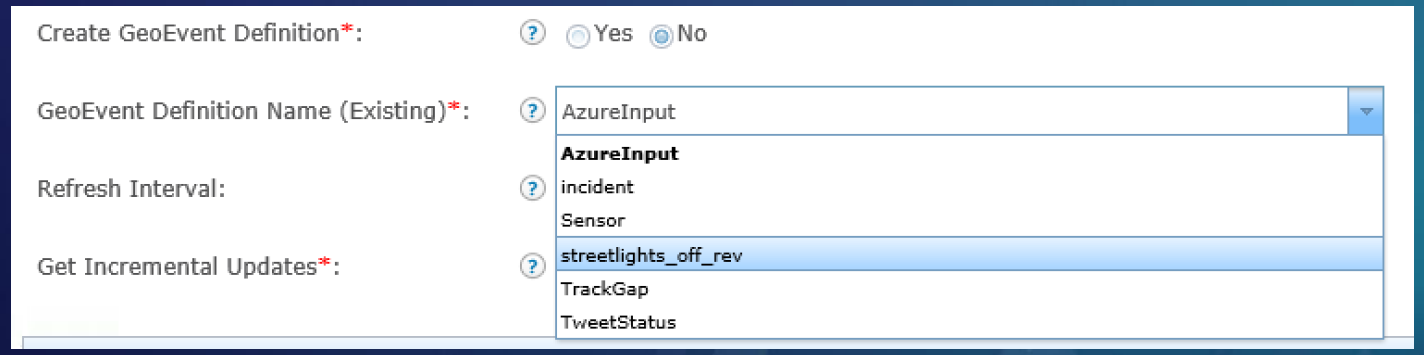
Troubleshooting

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

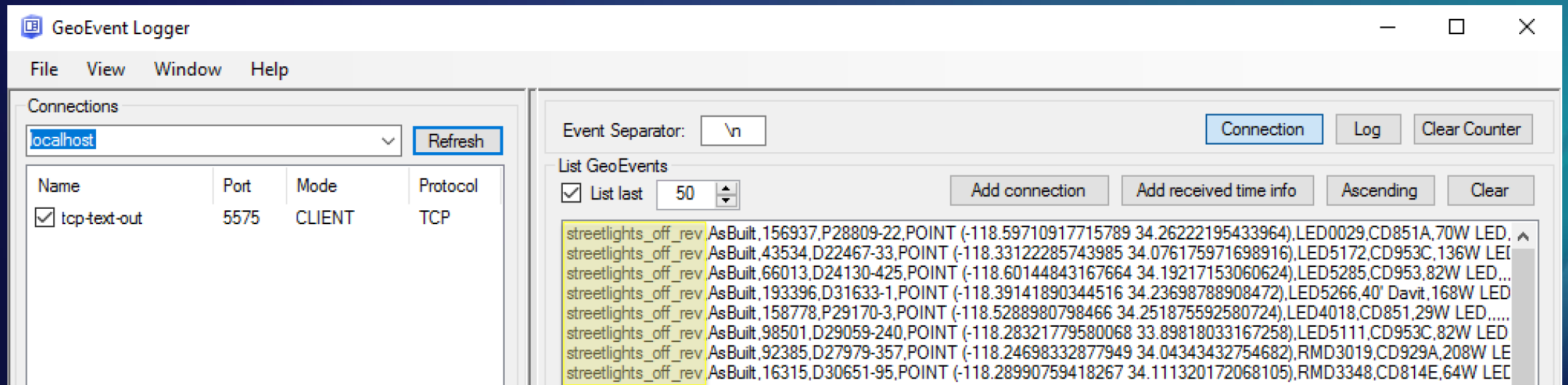


Troubleshooting

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…

Name	Value	Action
Automatic Backup Settings		
Backup Options	Back-up type: Daily; Properties: Time: 00:00	
Enable Automatic Backups	true	
Max Backups	10	

Edit Folder

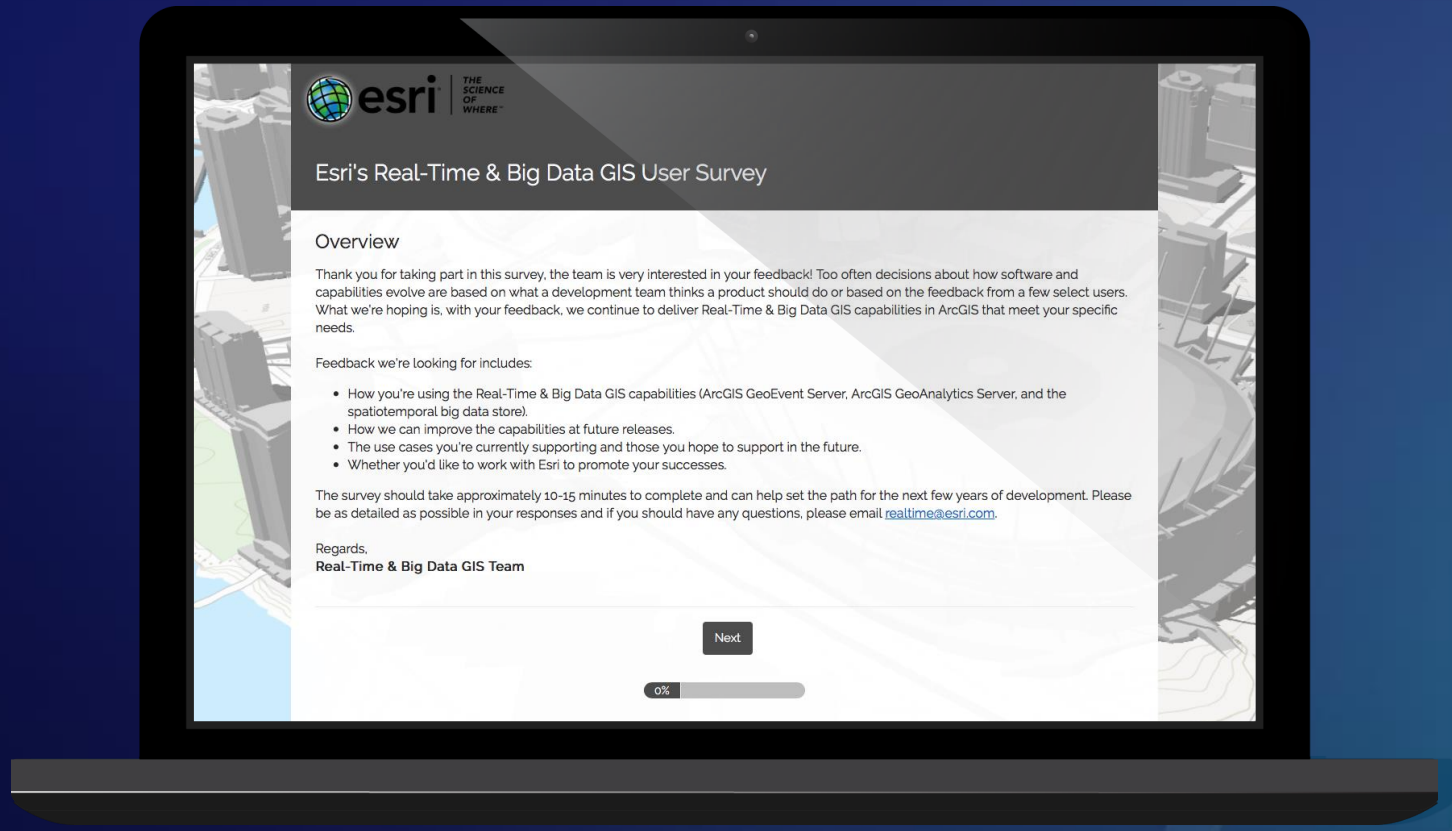
Register a folder.

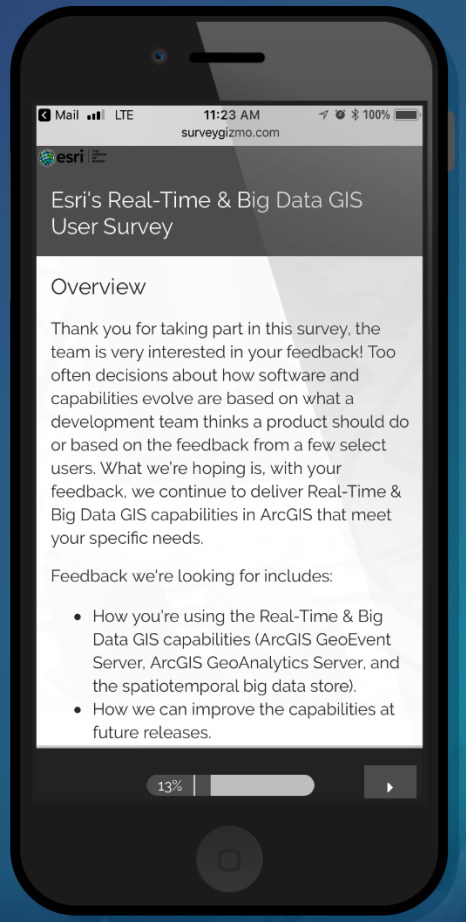
Name: Automatic Backups

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

esri | THE SCIENCE OF WHERE