Real-Time & Big Data GIS: Best Practices

Josh Joyner

RJ Sunderman

SEE WHAT OTHERS CAN'T

ArcGIS Enterprise

With Real-Time Capabilities

ingestion

actuation

Apps

Desktop

APIs

Live Features
Stream Services

Live & Historic
Aggregates & Features

visualization

ArcGIS Enterprise

ArcGIS GeoEvent
Server Role

analytics

Spatiotemporal
Big Data Store

storage

Agenda

1. Architecture Recommendations
2. Big Data Storage
3. Resiliency & Scalability
4. Stream Services
5. Service Design Considerations
6. Things I Wish I Knew

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 GB/s*
- **Memory**
  - **size – 8GB has been required since 10.3.** *32GB, default JVM max heap size is 4 GB*
  - **type – minimum of DDR4 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 (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

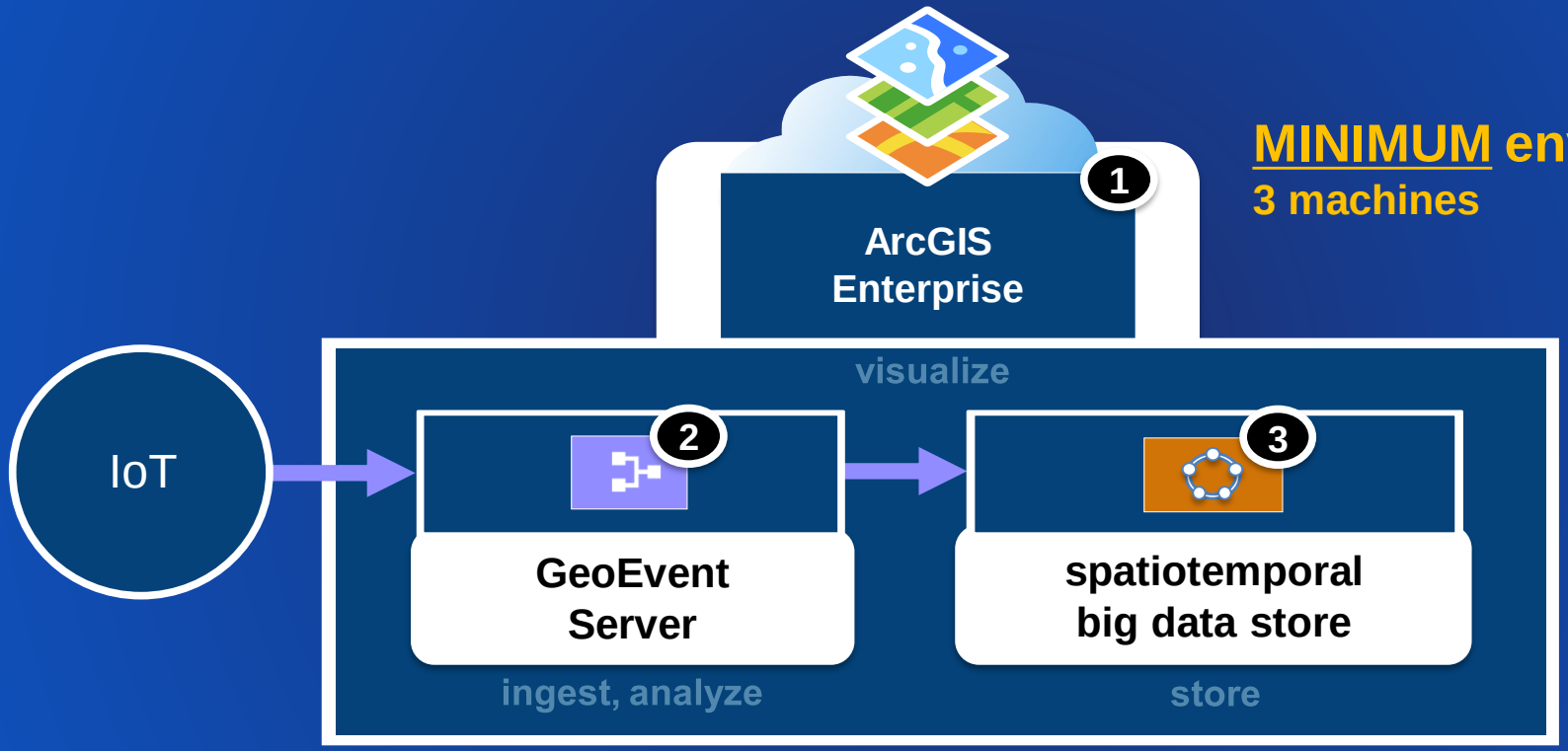
spatiotemporal big data store

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 GB/s EBS, note: local SSD is much better***
- **Network**
  - **speed – the faster the better.** ***1 GB/s***
- **Memory**
  - **size – 16GB minimum.** ***32GB, big data store allocates 8GiB by default***
  - **type – DDR4 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

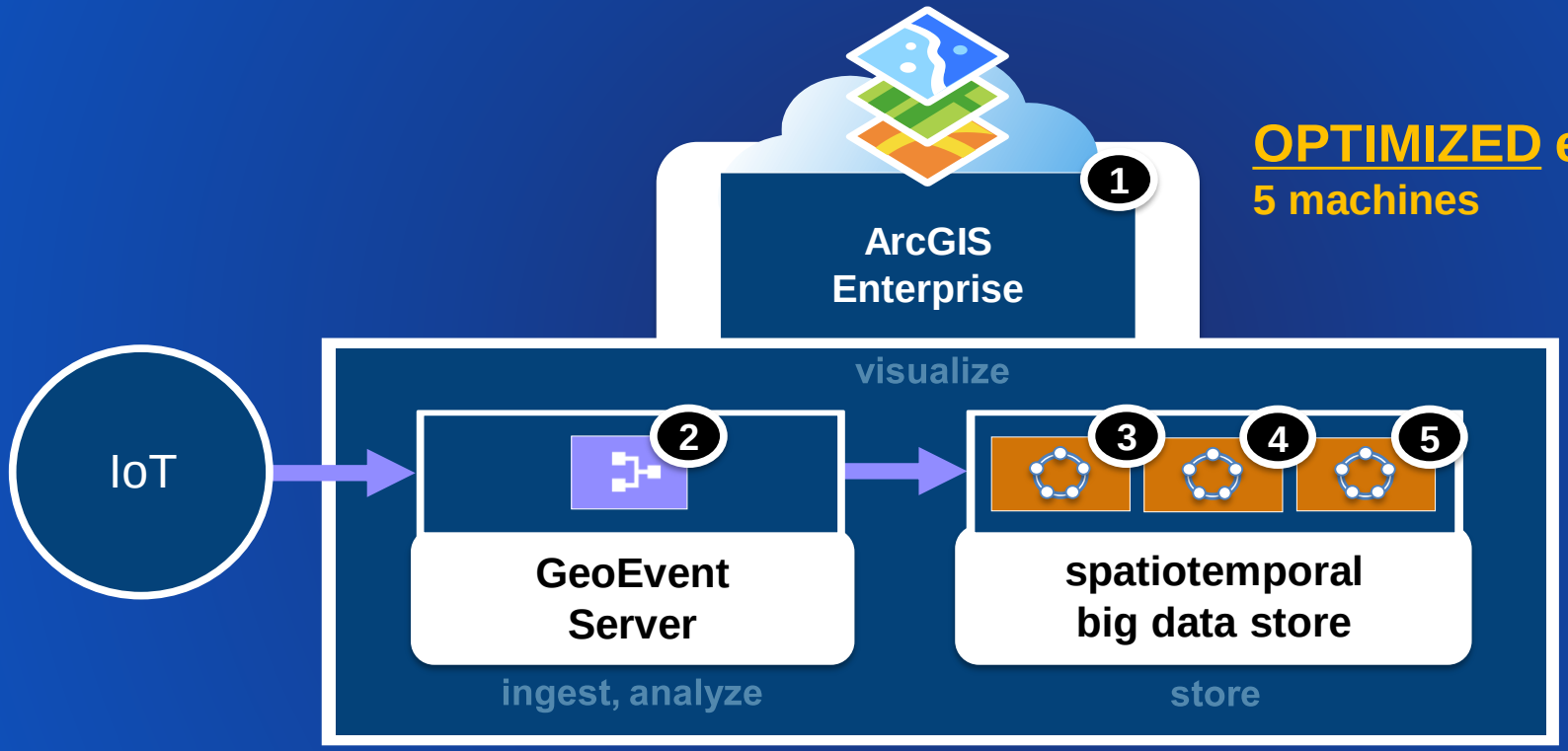
3 machines

functional servers & spatiotemporal big data store

SHOULD BE on ISOLATED machines

ArcGIS Enterprise

with real-time capabilities



OPTIMIZED environment

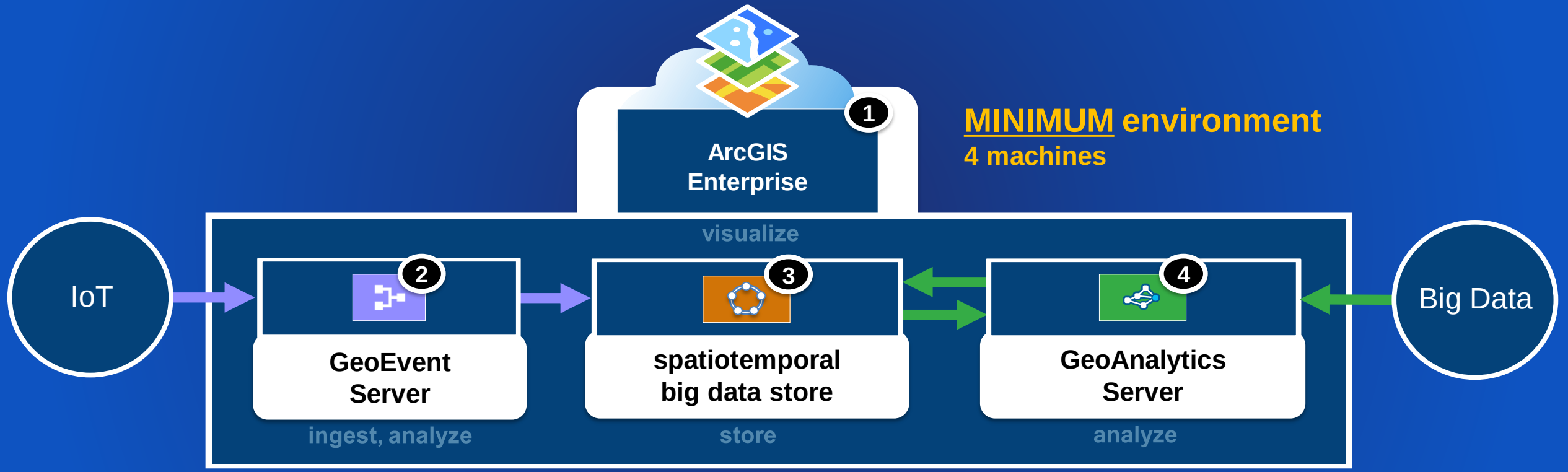
5 machines

functional servers & spatiotemporal big data store

SHOULD BE on ISOLATED machines

ArcGIS Enterprise

with real-time & big data GIS capabilities

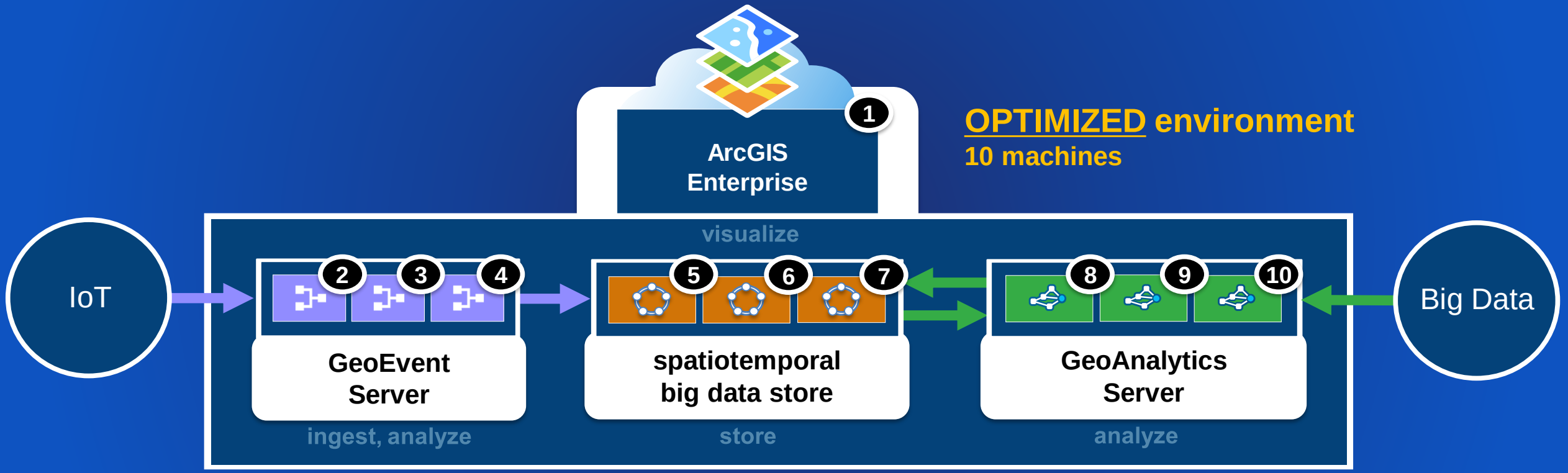


MINIMUM environment

4 machines

ArcGIS Enterprise

with real-time & big data GIS capabilities



GeoEvent Server

updating Gateway deployment location

- **GeoEvent Gateway is deployed to**

  “C:/ProgramData/Esri/GeoEvent-Gateway/” (Windows OS)

  - If utilizing a large number of inputs or outputs this location can be changed by editing:

  - (ArcGIS Server Installation Path)\GeoEvent\gateway\etc\kafka.properties

```
gateway.data.dir=C://ProgramData//Esri//GeoEvent-Gateway//
```

  - (ArcGIS Server Installation Path)\GeoEvent\gateway\etc\zookeeper.properties

```
gateway.data.dir=C://ProgramData//Esri//GeoEvent-Gateway//
```

GeoEvent Server

other deployment considerations

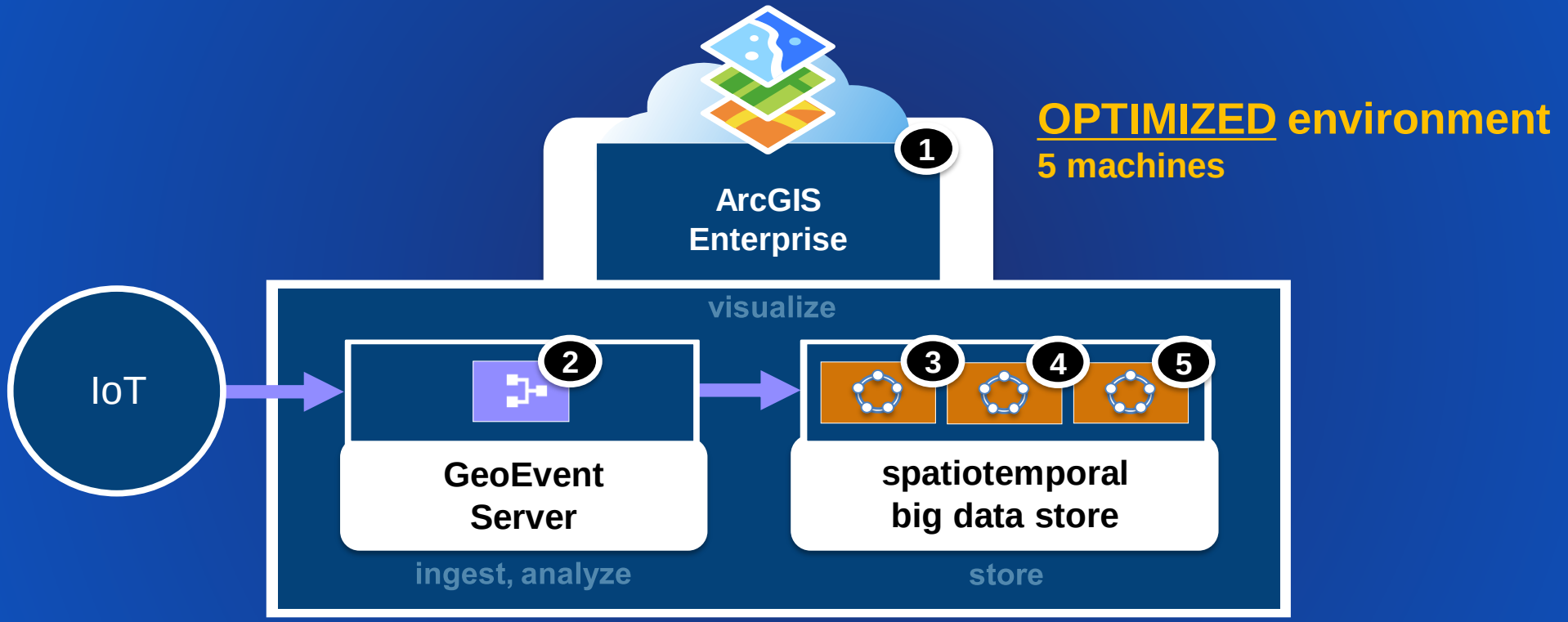
- **GeoEvent Server and GeoEvent Gateway services should be left running**
  - Limit system restarts to scheduled patch times
  - Inputs and Outputs leveraging Hosted Feature Layers should be stopped prior to any required restarts
  - Be aware of server and service dependencies

- **Patches and Upgrades typically require the reset/clearing of browser cache to complete installation**

2 Big Data Storage

ArcGIS Enterprise

with real-time capabilities

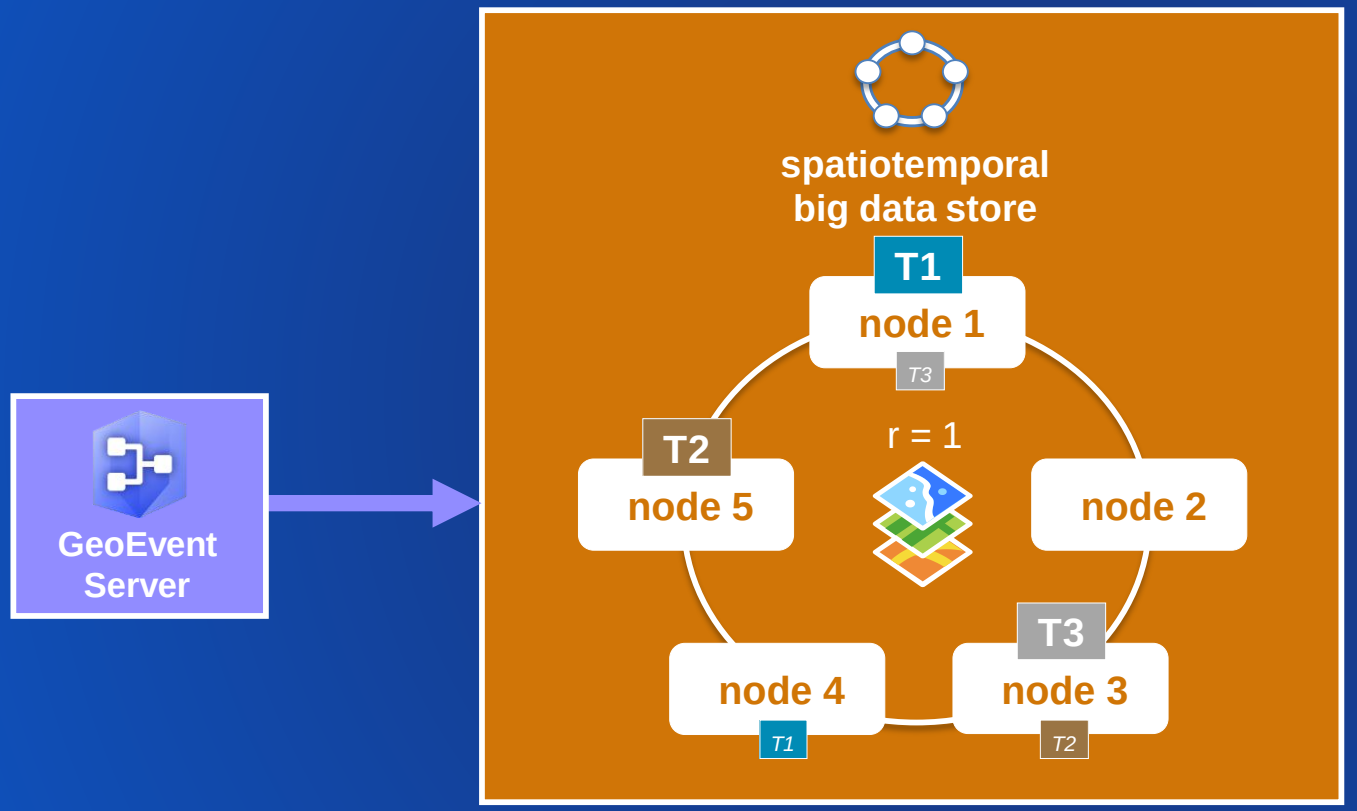


OPTIMIZED environment

5 machines

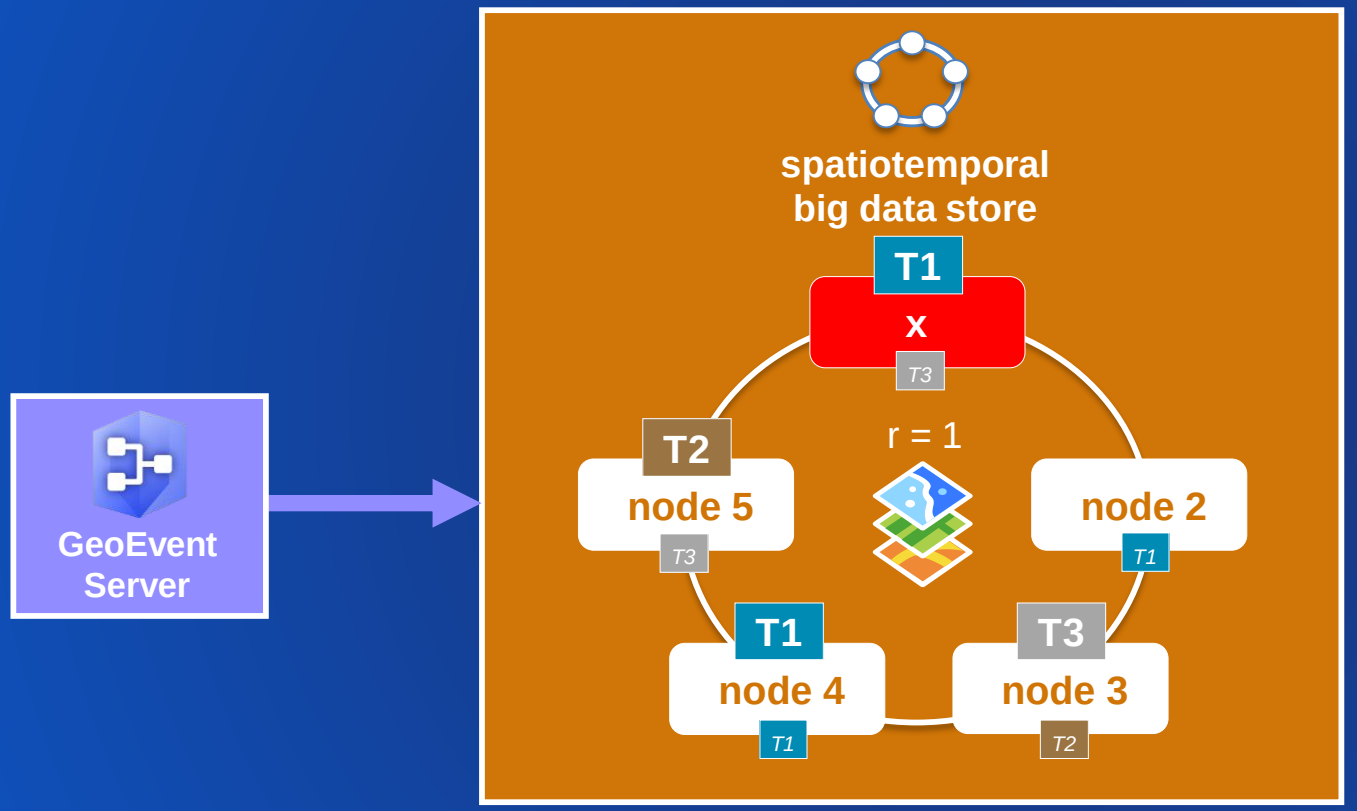
spatiotemporal big data store

shards & replication factor



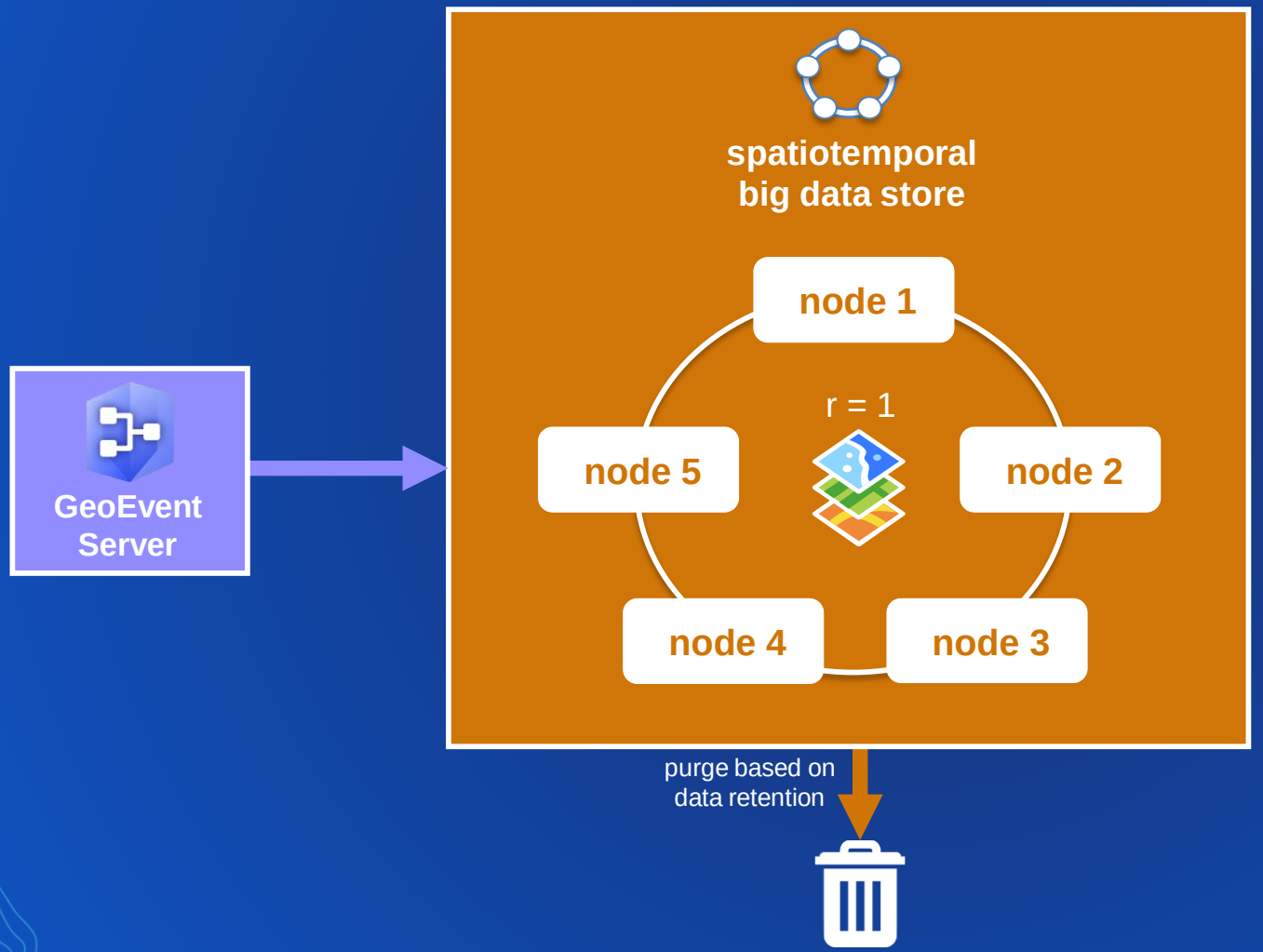
spatiotemporal big data store

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



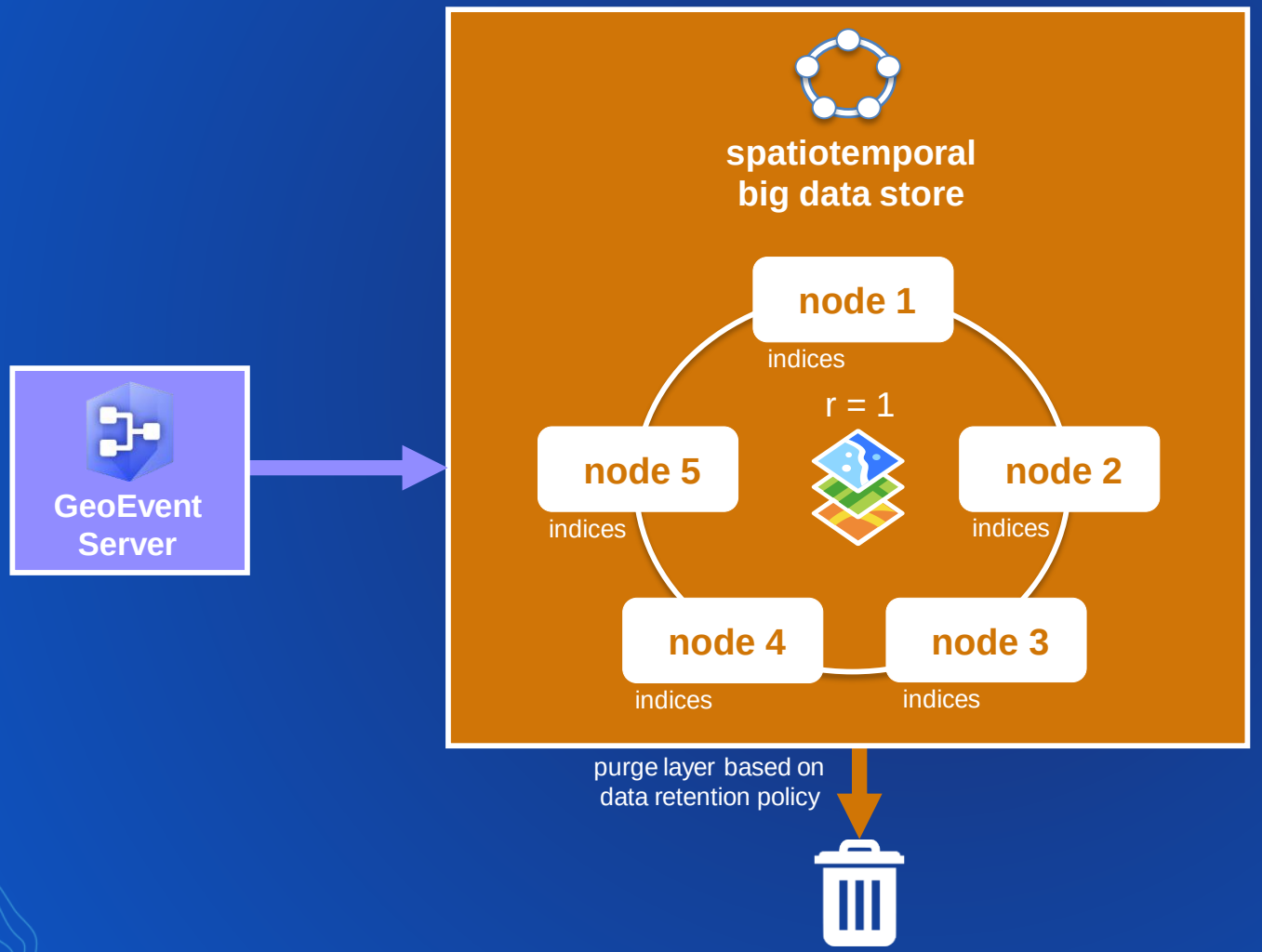
spatiotemporal big data store

data retention policies, configured per data source



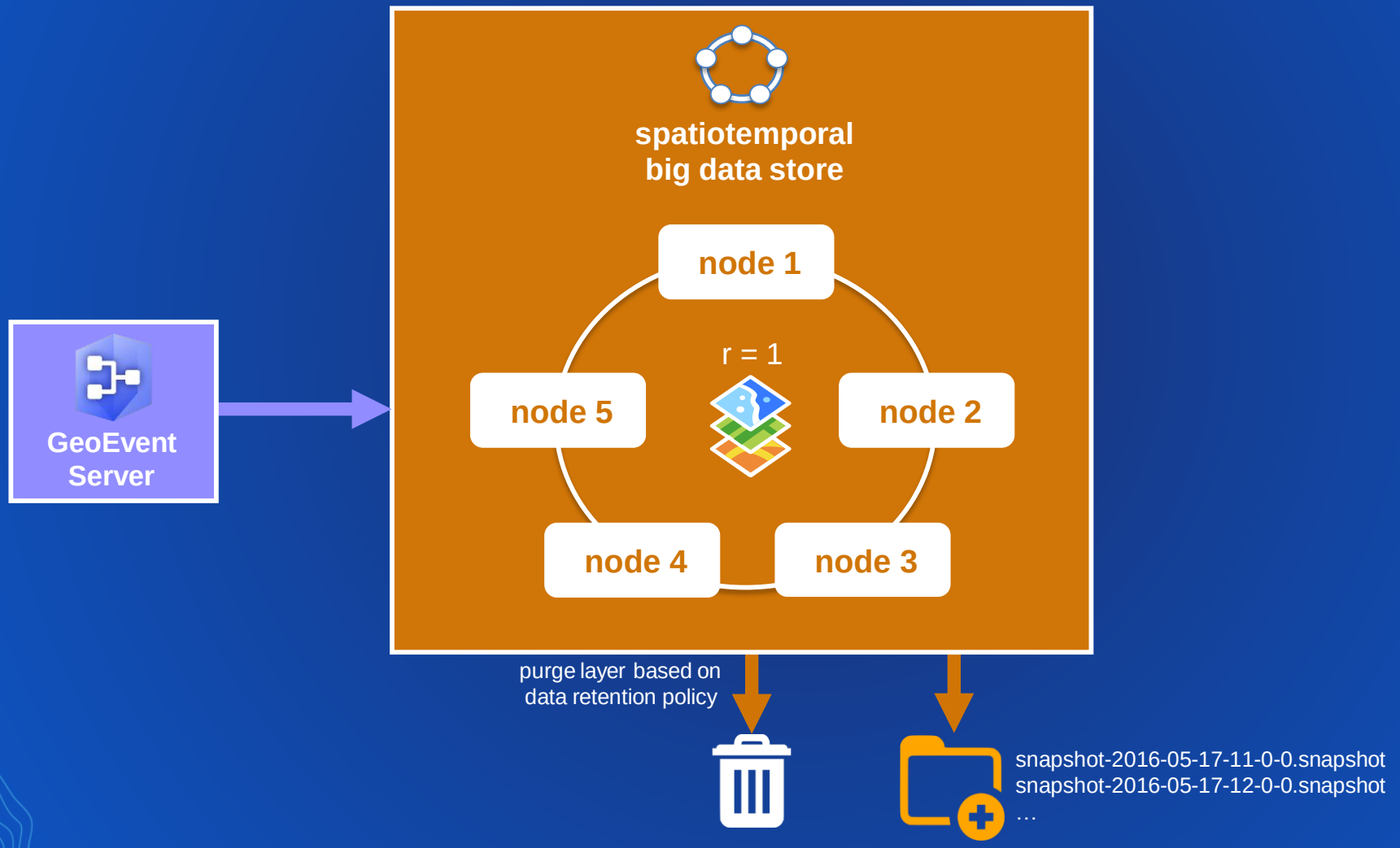
spatiotemporal big data store

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



spatiotemporal big data store

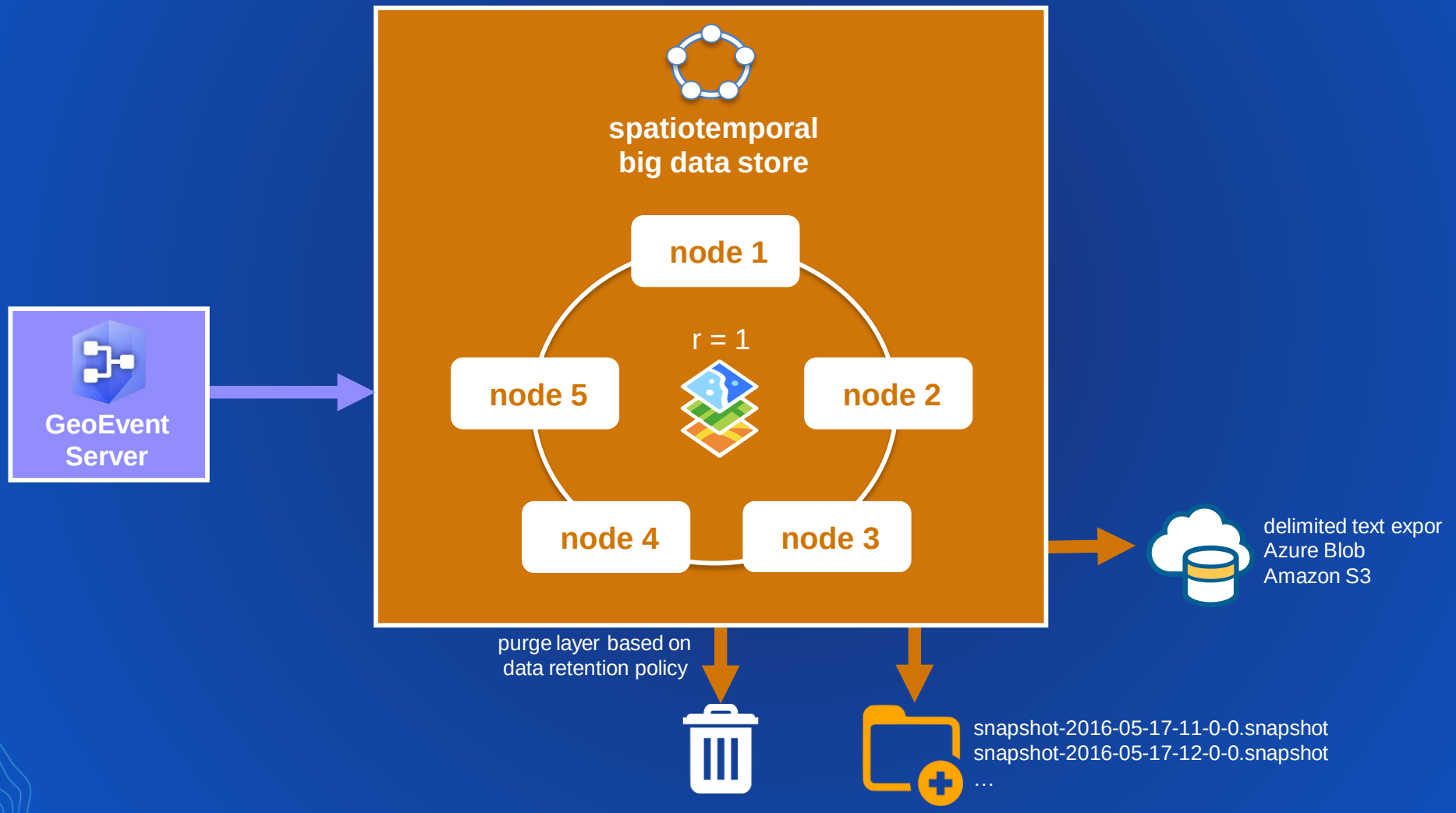
automatic data backups using periodic snapshots, including ability to restore from a snapshot



spatiotemporal big data store

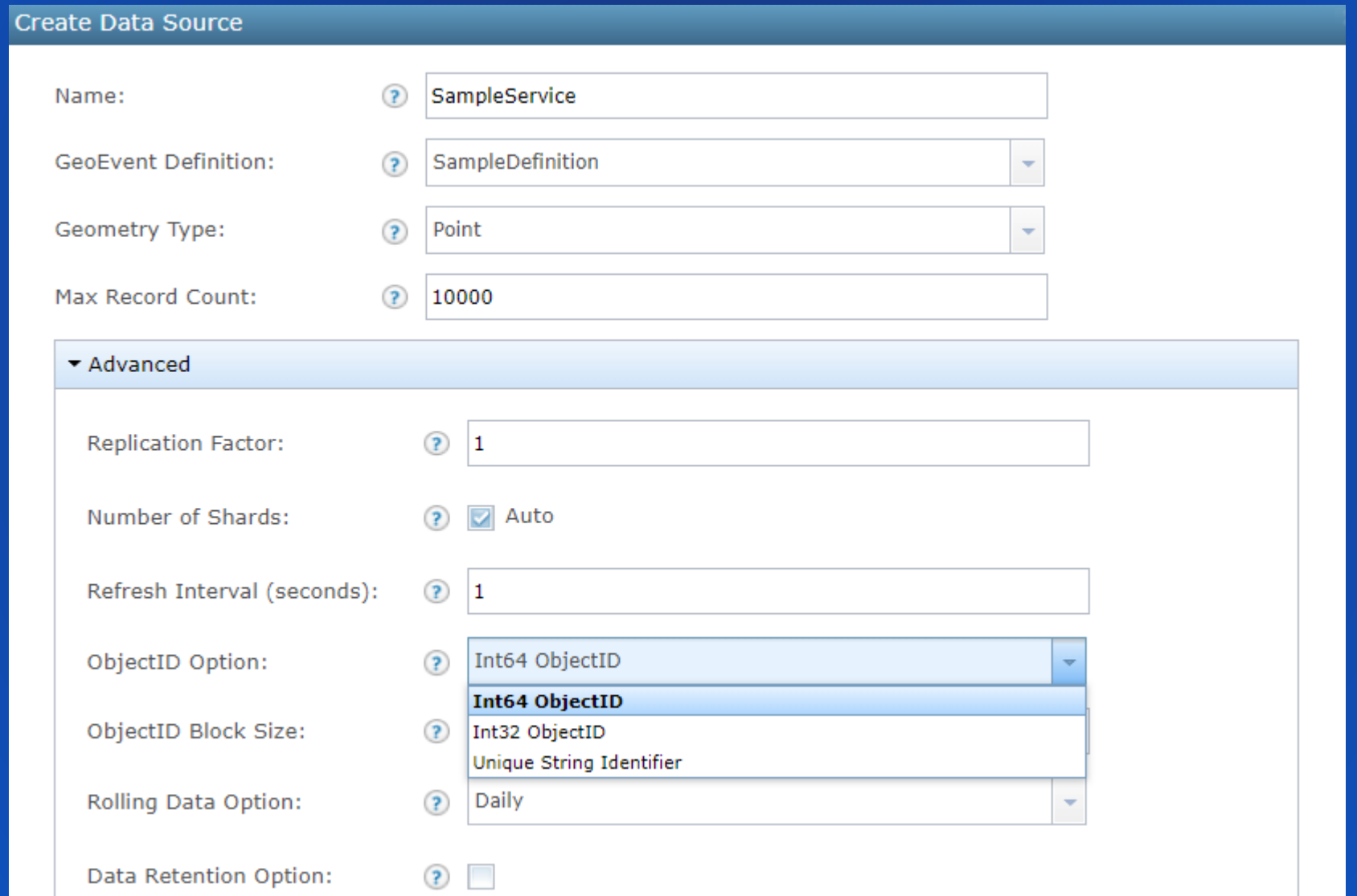
manual and programmatic export data to cloud stores

10.7



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

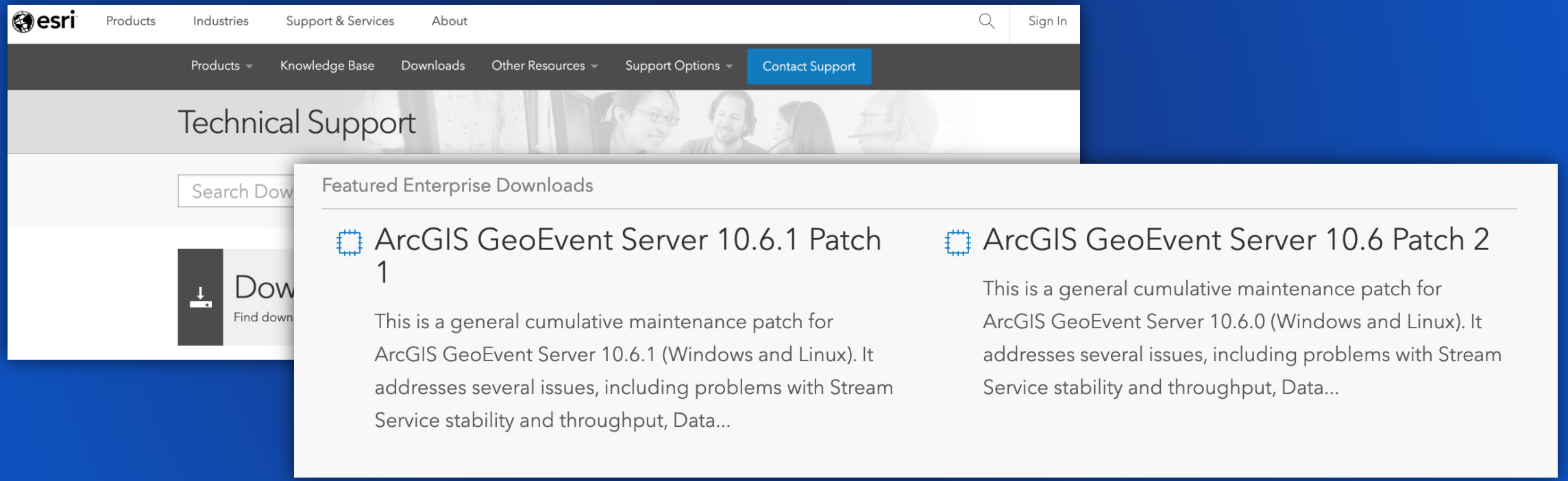
IoT rate *per second*	per day	per week	per month	per year
1	86,400	604,800	2,592,000	31,104,000
10	864,000	6,048,000	25,920,000	311,040,000
100	8,640,000	60,480,000	259,200,000	3,110,400,000
1,000	86,400,000	604,800,000	2,592,000,000	31,104,000,000
10,000	864,000,000	6,048,000,000	25,920,000,000	311,040,000,000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

stay current with releases

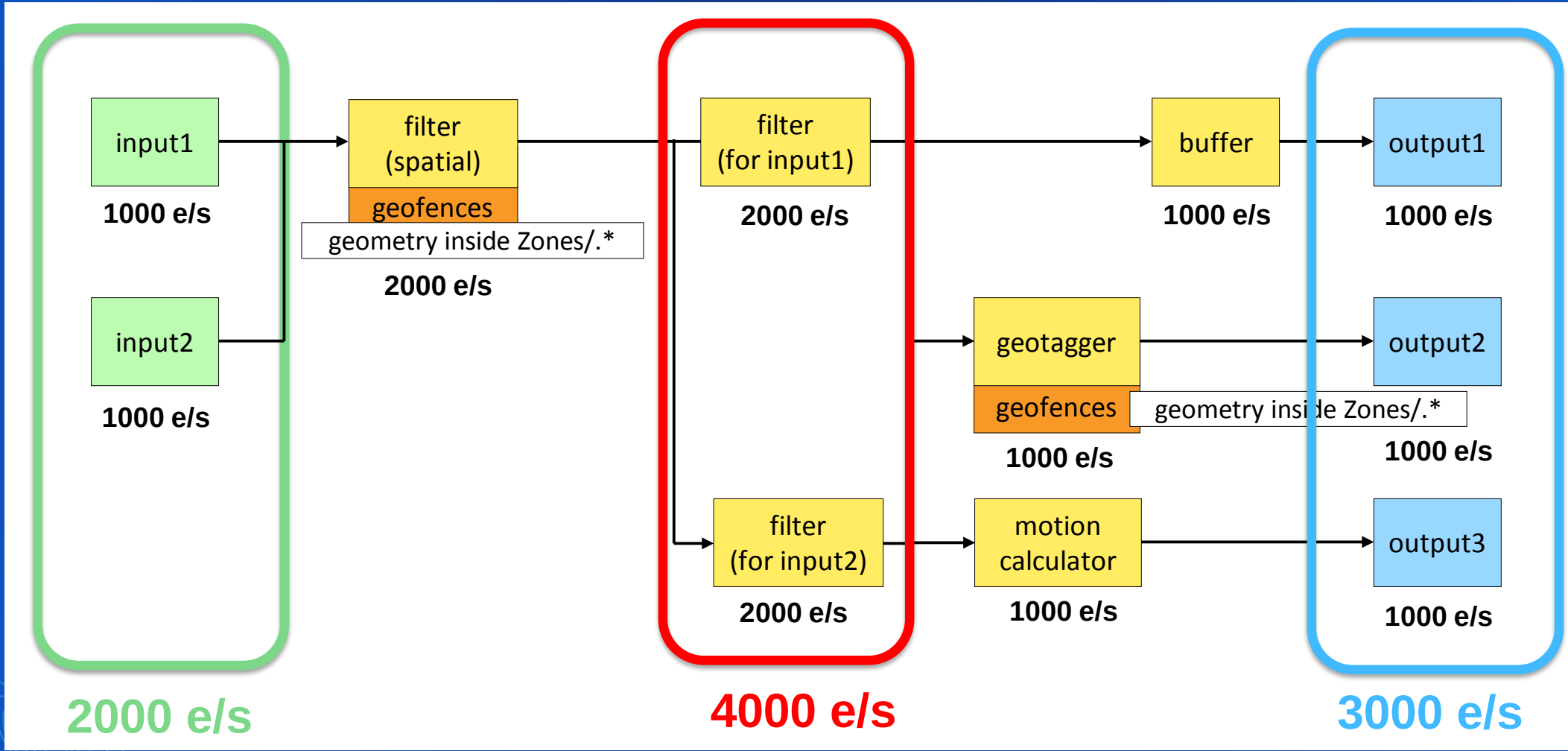
- **New Patches available for 10.6.0 & 10.6.1 releases**



https://support.esri.com/en/downloads

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)

```
# Minimum and Maximum Java Heap Sizes
wrapper.java.additional.13=-Xms1g
wrapper.java.additional.14=-Xmx4g
```

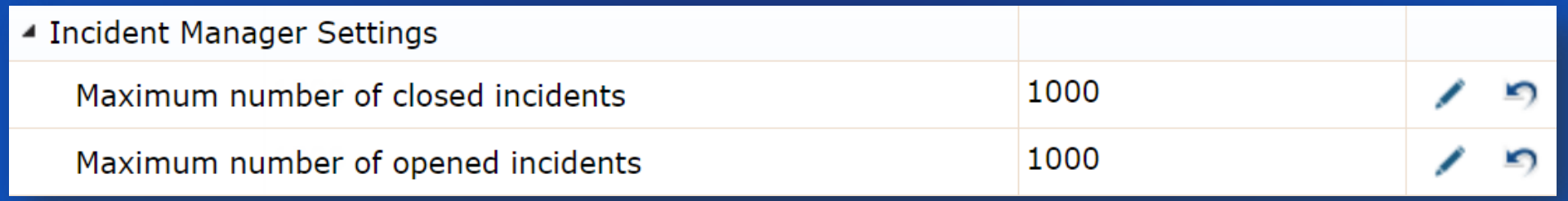
Performance, Resiliency & Scalability

configuration changes to support larger scale

- Individual filters and processors will cache up to 1000 unique Track_ID values by default.
  - 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



4 Stream Services

Stream Services

why use a stream service instead of a feature service?

- Stream Services allow for improved rendering detail of moving objects
  - In a Web Map / App updates can be rendered every 100 milliseconds
  - Feature layers poll for updates once every 6 seconds
- Light weight requiring no storage of events
- Updates are pushed instead of polled, supporting reduced delays
- With Pro 2.2+ users are able to configure unique feature rendering

Stream Services

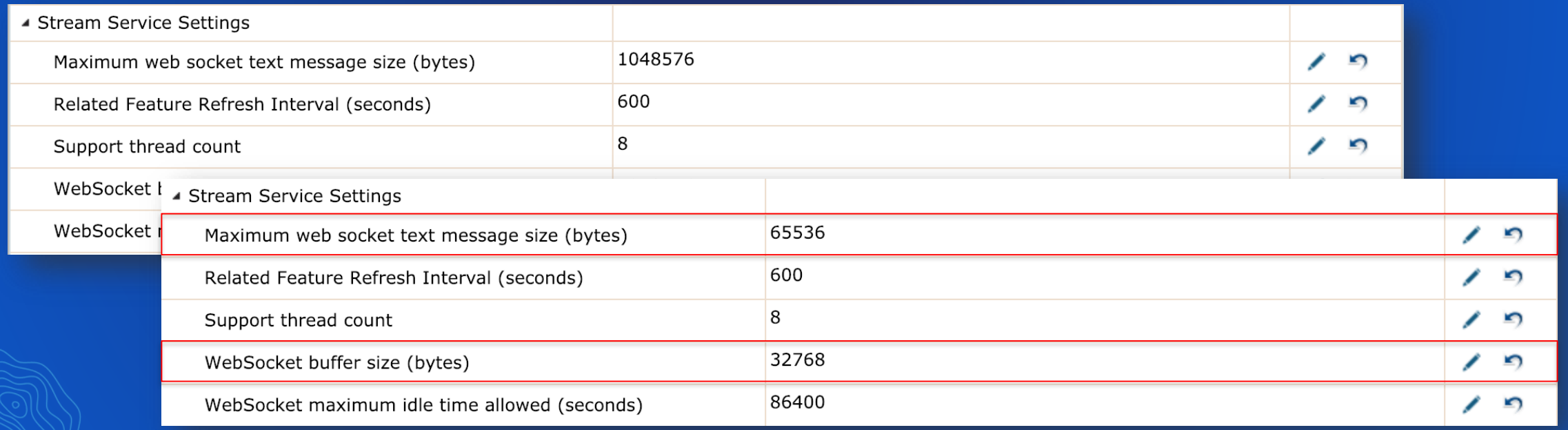
why use a feature service instead of a stream service?

- Feature Services provide a confirmation that add/updates have been completed
  - Stream Services do not provide any delivery response (greater risk of missing messages)
- Unique symbology support across all web and desktop clients
  - Aggregation support with the spatiotemporal big data store
- Data is persisted even when client is not actively receiving data
  - Stream Service messages are ephemeral, being lost when no client is subscribed

Stream Services

other considerations

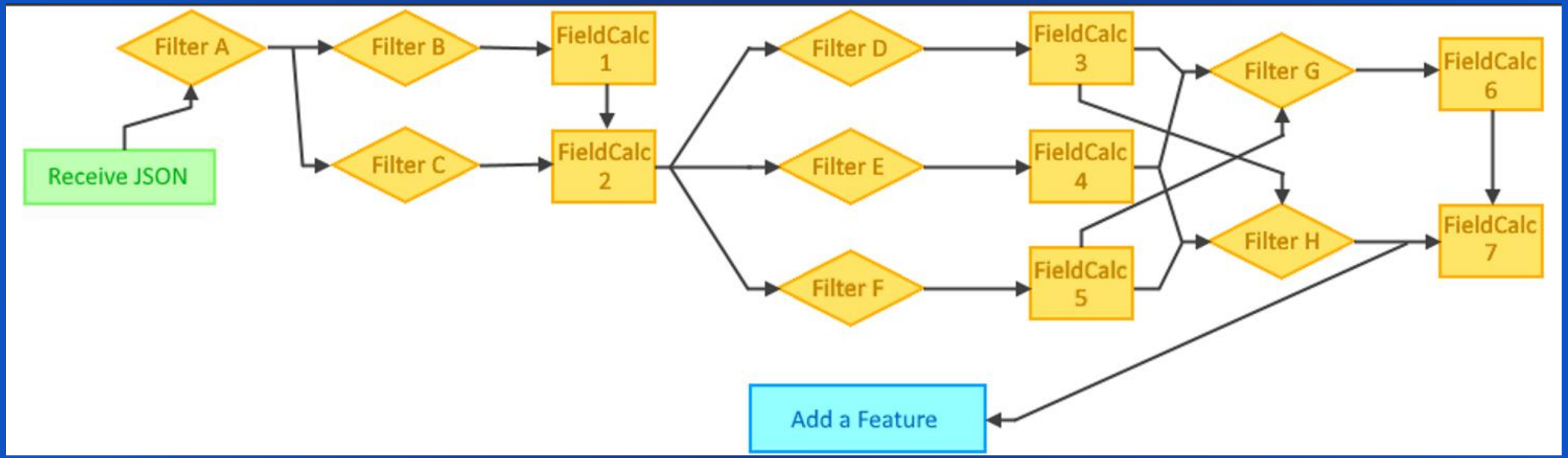
- Can you have both?
  - Configure a Stream Service with Store Latest
  - Bootstrap records from a feature service and replaced with new data from the stream
- Adjust Global Settings for improved throughput



5 Service Design Considerations

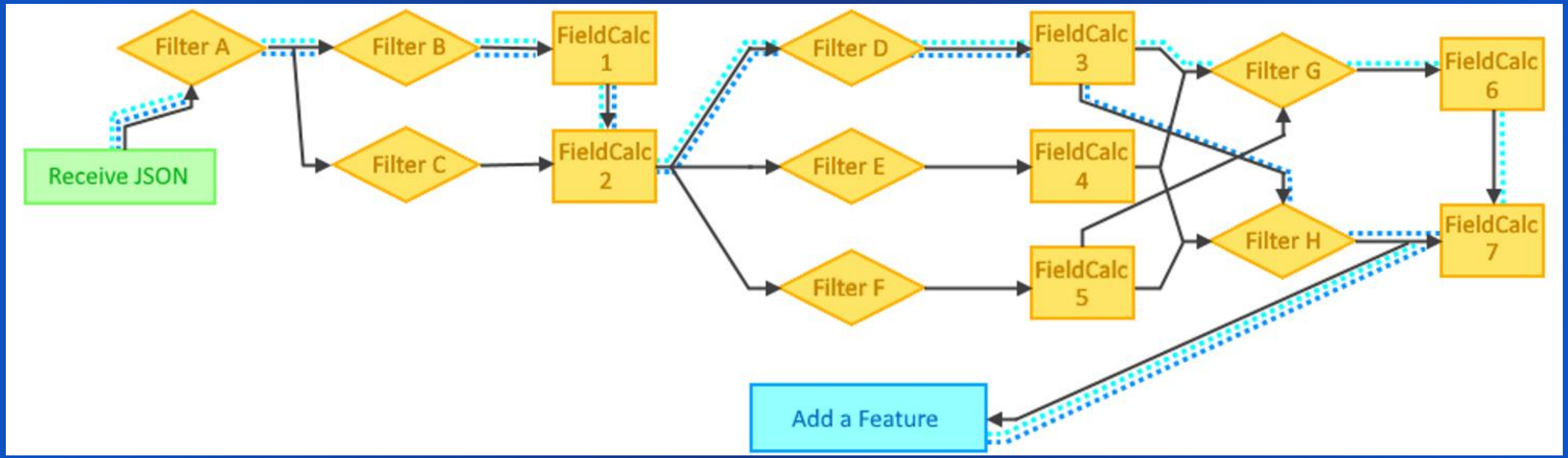
Service Design Considerations

how many event processing paths do you see?



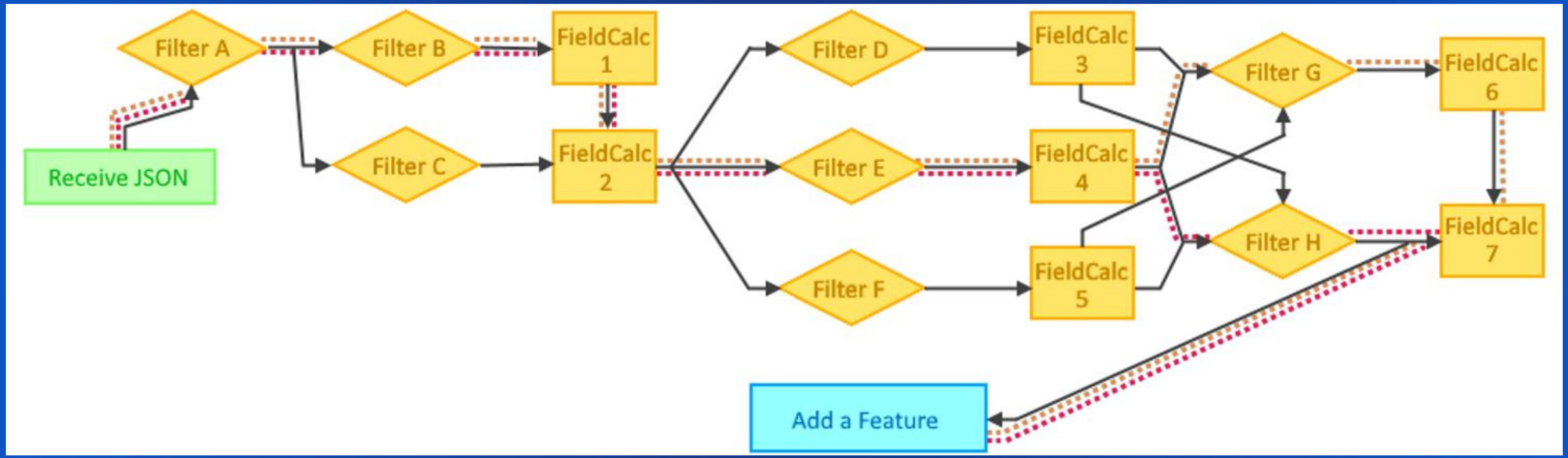
Service Design Considerations

how many event processing paths do you see?



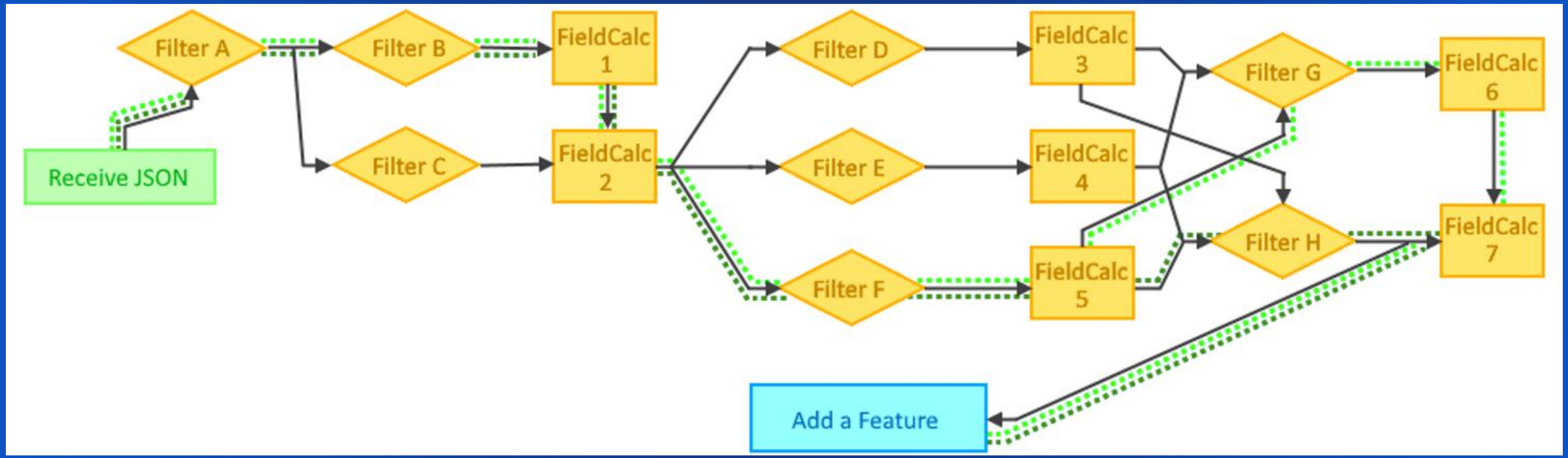
Service Design Considerations

how many event processing paths do you see?



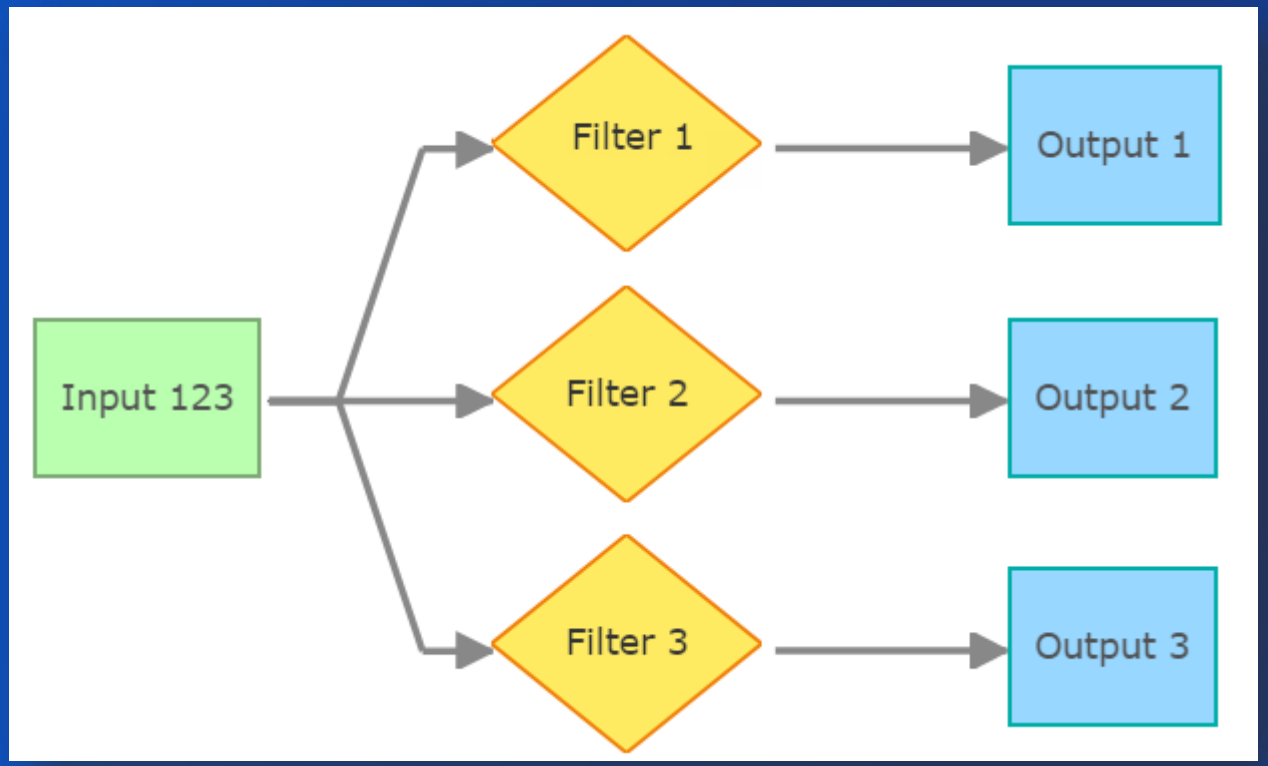
Service Design Considerations

how many event processing paths do you see?

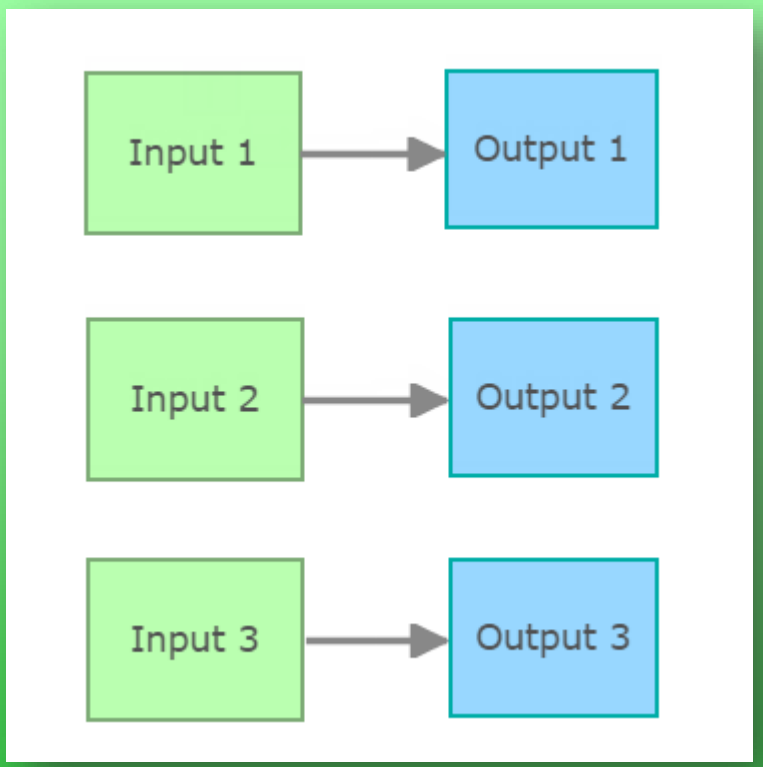


Service Design Considerations

which would you choose?



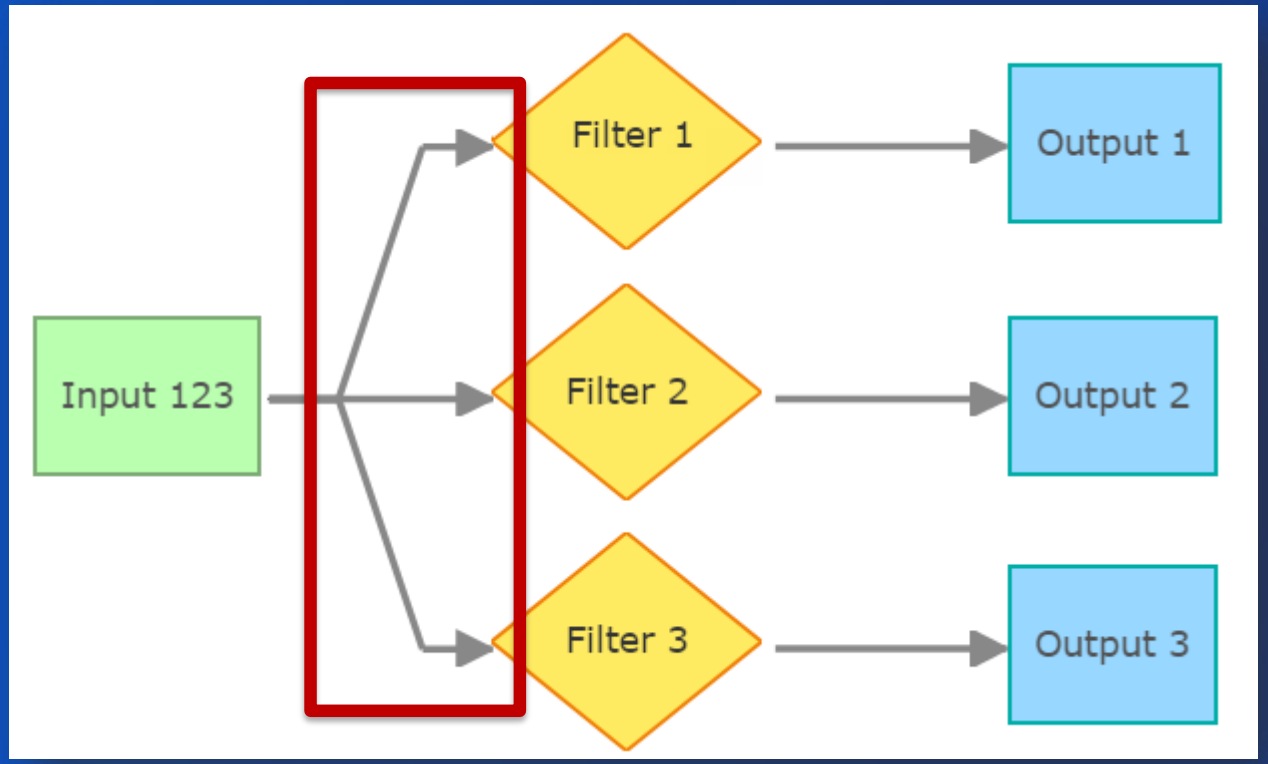
Service A



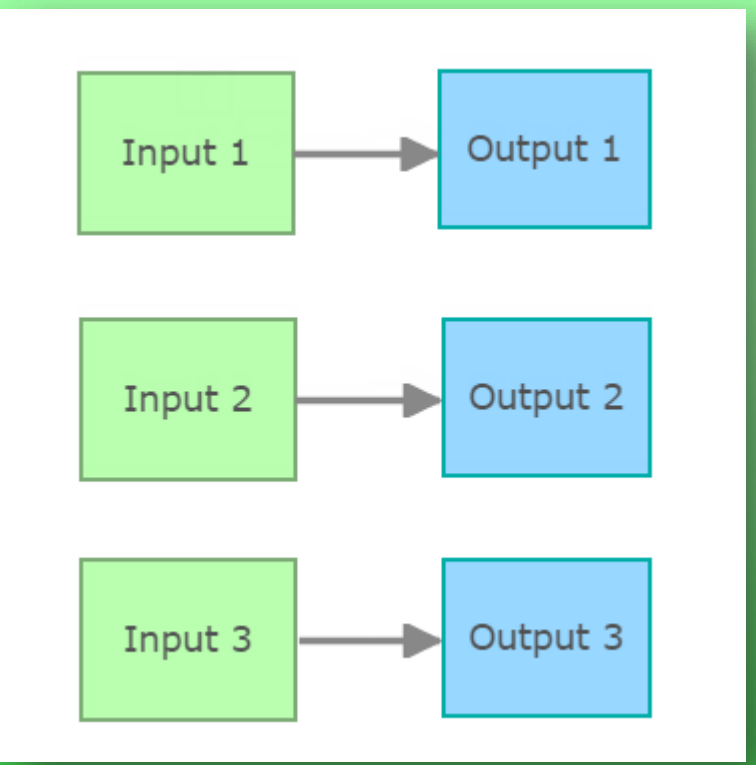
Service B

Service Design Considerations

which would you choose?



Each “branch” in a service contains the same event data. In this example the three branches triple the number of records the underlaying message bus needs to handle.

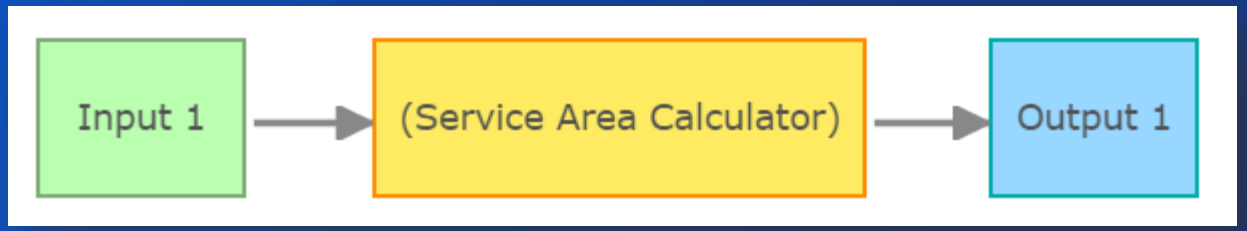


When possible, pre-filter the input data before ingesting.

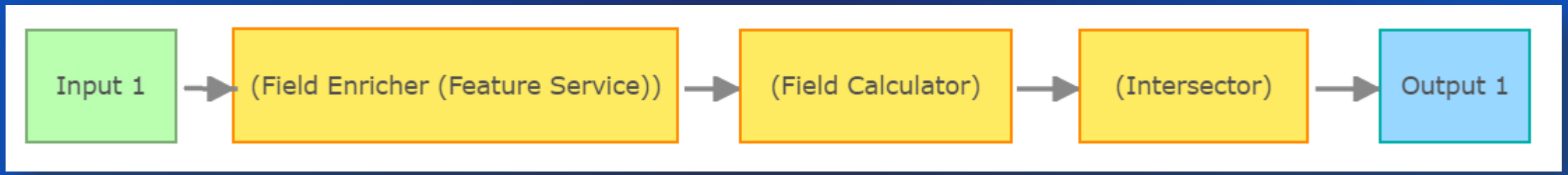
Service Design Considerations

not all components are created equally

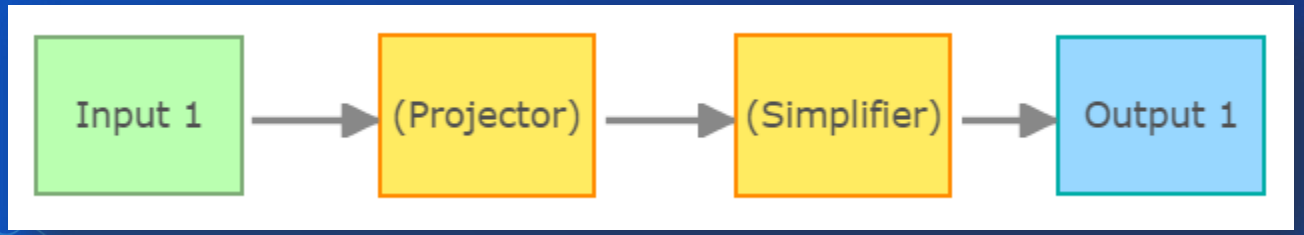
A



B



C

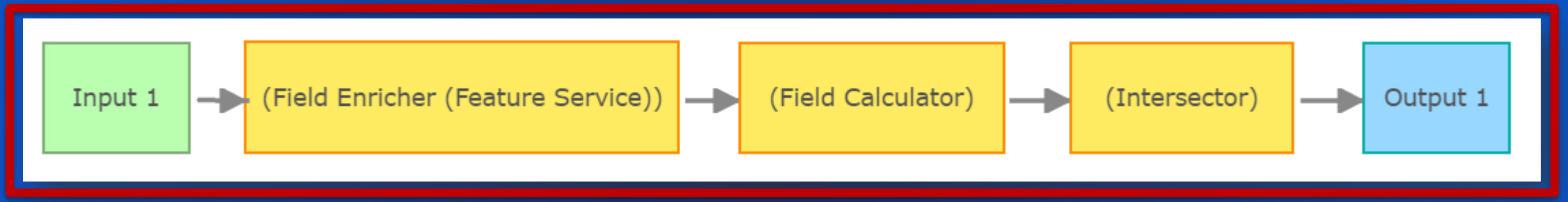


Which of these services has the potential to least impact performance?

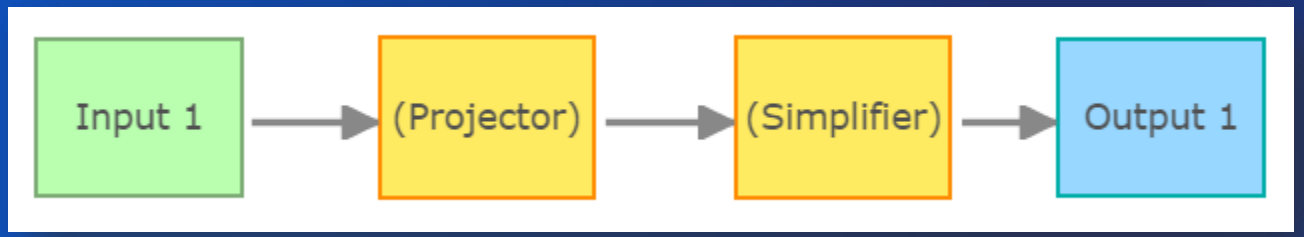
Service Design Considerations

not all components are created equally

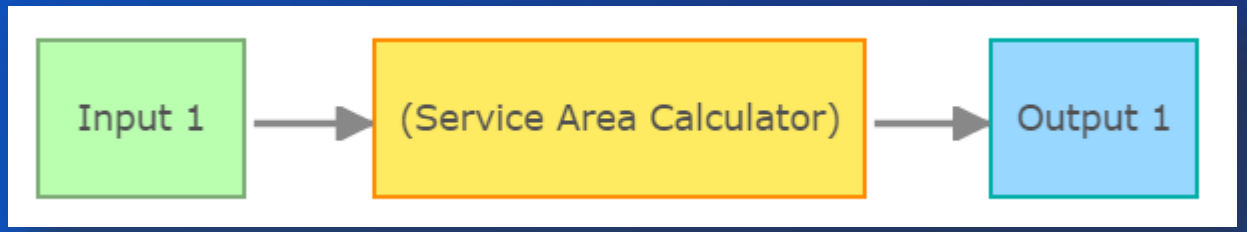
B



C



A

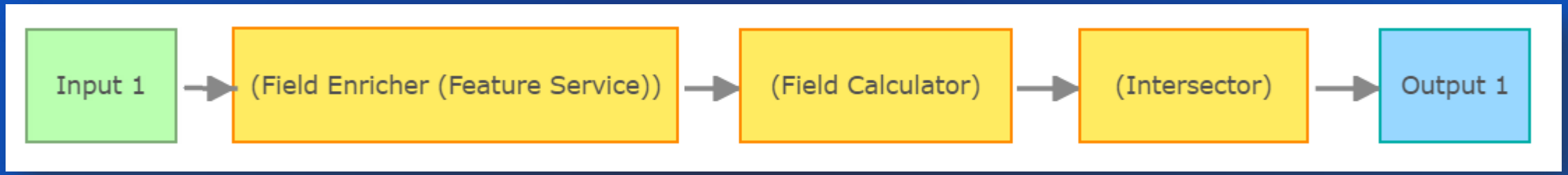


The first service, though it has more nodes, operates against a cache of enrichment records and in-memory geofences allowing for better performance.

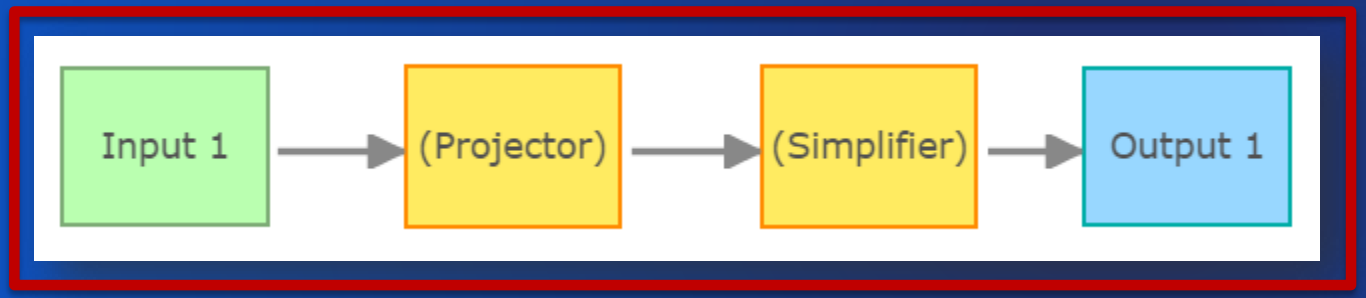
Service Design Considerations

not all components are created equally

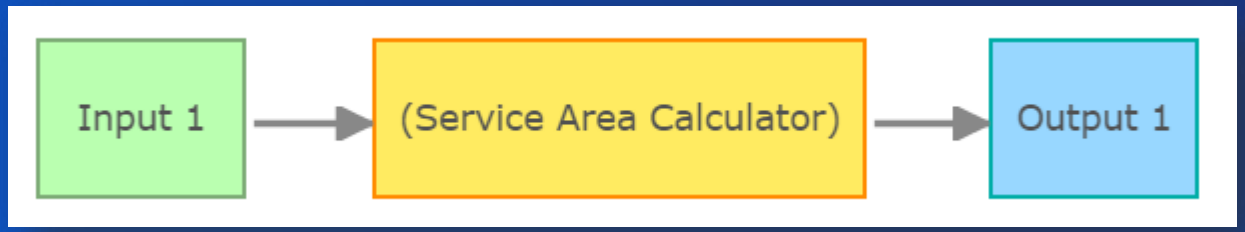
B



C



A



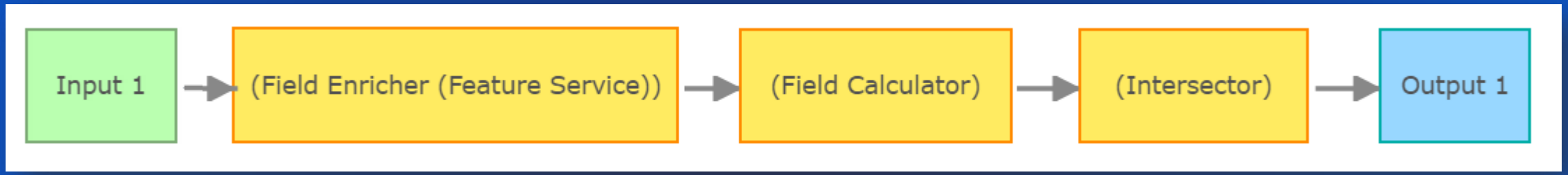
The second service modifies the incoming event geometry which can be “expensive”.

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

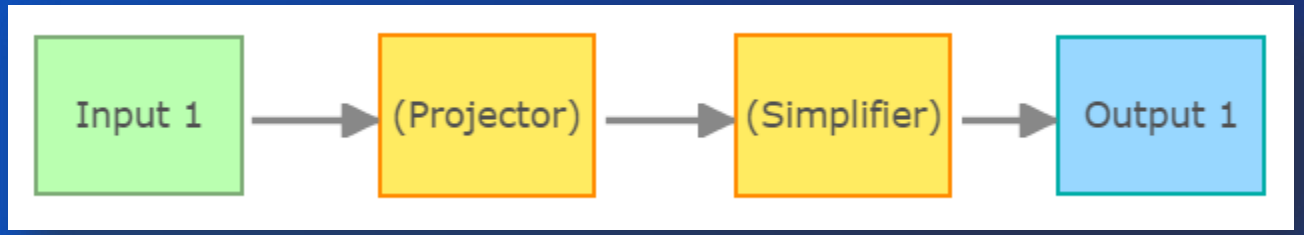
Service Design Considerations

not all components are created equally

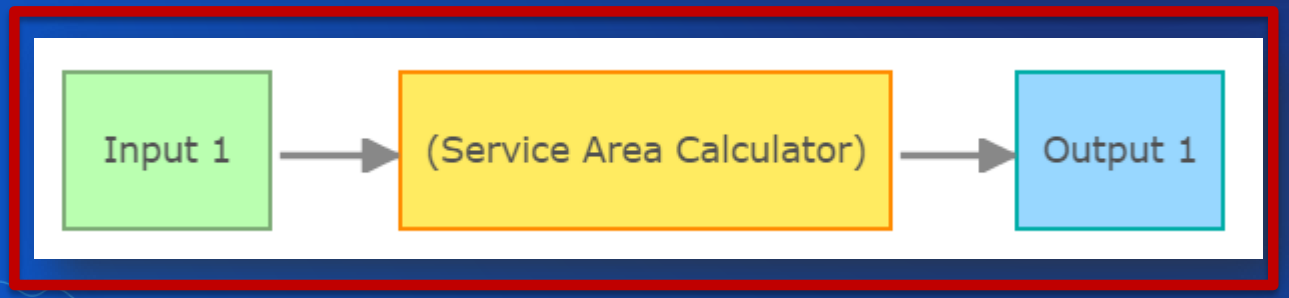
B



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 Field Mapper 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 single data pipelines whenever possible**
  - Branching out or back together of pipelines can have multiplicative impacts
  - Better to duplicate workflows across multiple services than a single complex workflow

6 Things I Wish I Knew

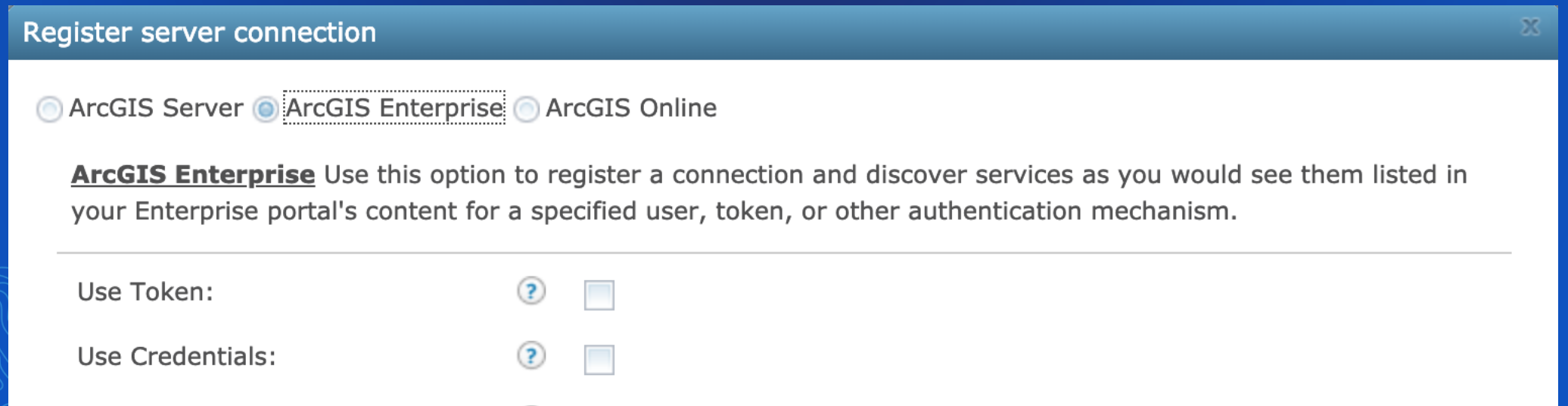
Things I Wish I Knew

federating GeoEvent Server

"I have to federate the GeoEvent Server… don't I?"

- In order to use the spatiotemporal big data store it requires an Enterprise type connection

- Federating the local ArcGIS Server (and in turn GeoEvent Server) provides…
  - Single Sign-on Experience
  - Automatically converting Default datastore type
    - …but even if you don't federate you can still make a Datastore Connection to your ArcGIS Enterprise

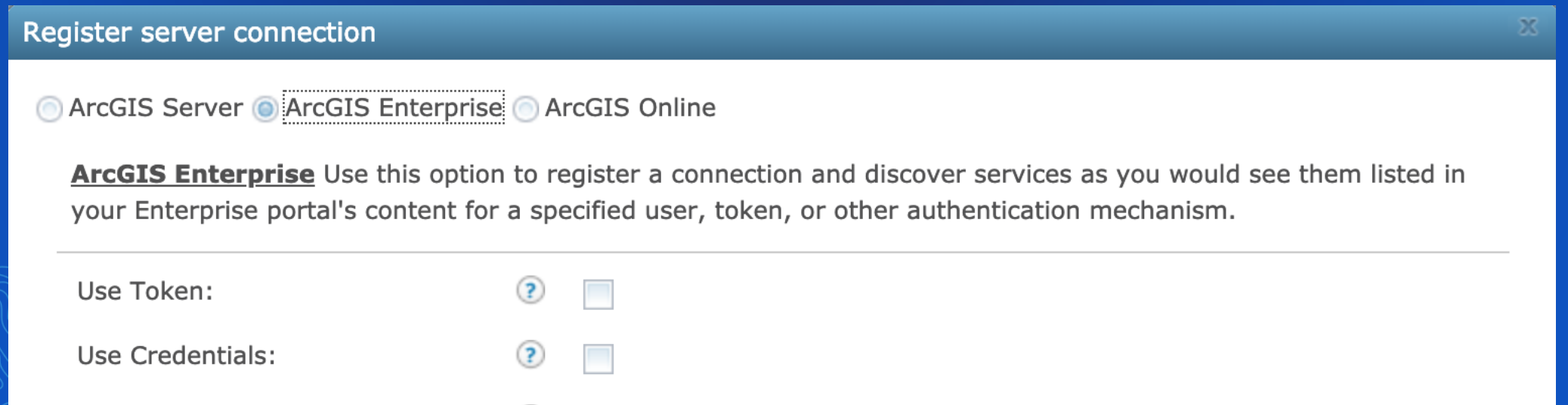


Things I Wish I Knew

federating GeoEvent Server

"I have to federate the GeoEvent Server… don't I?"

- Some deployments (especially in highly secured environments) function better without federating
- It also allows more restrictive access to the GeoEvent Manager and Services
- Only those Admin level users on the local ArcGIS Server can log into GeoEvent Manager

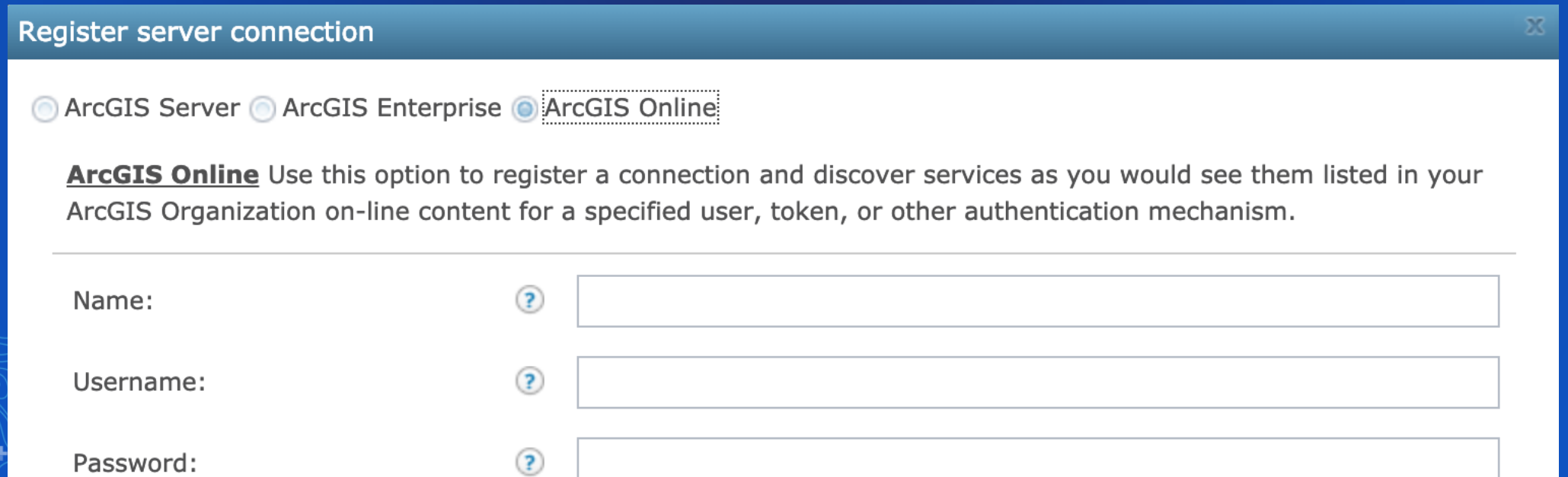


Things I Wish I Knew

registered data store connections

"What's the difference in the types of Registered server connections"

- 3 Types: ArcGIS Server, ArcGIS Enterprise, ArcGIS Online
  - ArcGIS Online: Access your Organization Content
    - Supports reading and writing from Hosted Layers
    - Publishing of new content not supported

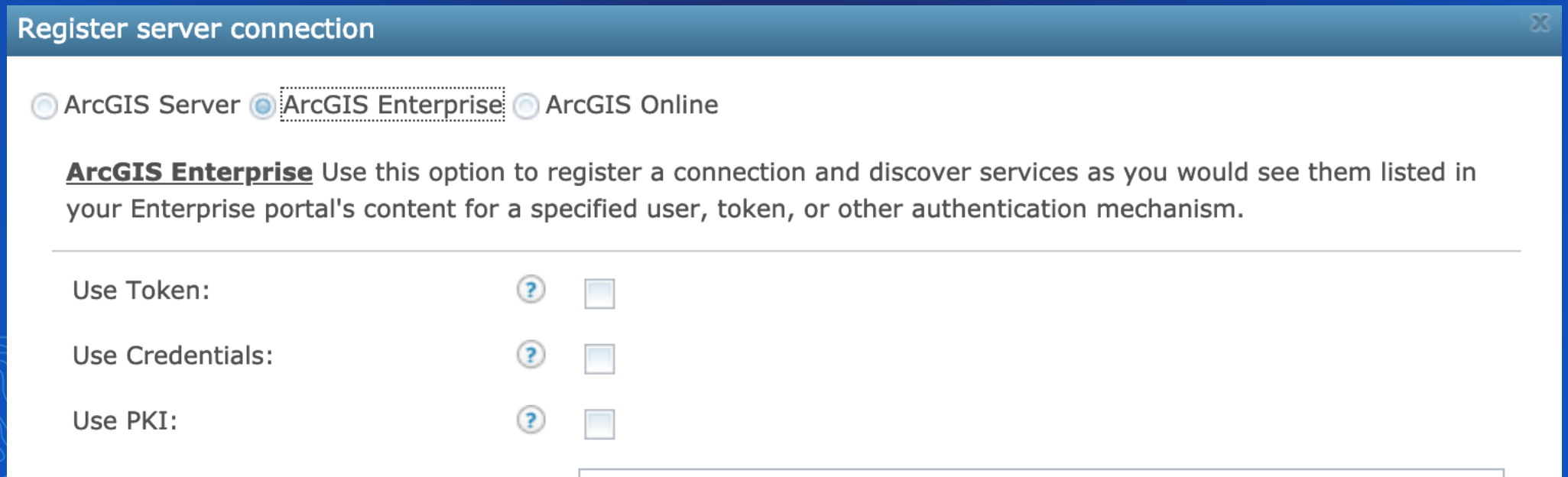


Things I Wish I Knew

registered data store connections

"What's the difference in the types of Registered server connections"

- 3 Types: ArcGIS Server, ArcGIS Enterprise, ArcGIS Online
  - ArcGIS Enterprise: User Level Access
    - Supports reading, writing, and publishing of Hosted Layers
    - Content access is limited to a defined set of credentials and based on item ownership

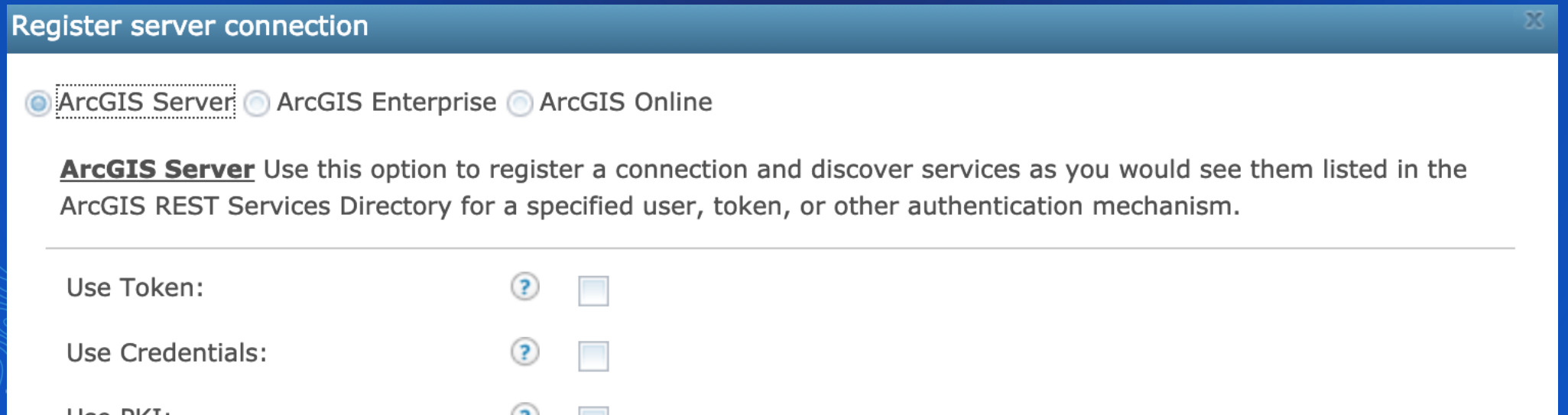


Things I Wish I Knew

registered data store connections

"What's the difference in the types of Registered server connections"

- 3 Types: ArcGIS Server, ArcGIS Enterprise, ArcGIS Online
  - ArcGIS Server: Role Level Access
    - Supports reading and writing of Hosted Service
    - Publishing restricted to Stream Services and Relational Feature Services
    - Content access is based on role, NOT the User



Things I Wish I Knew

version parity requirements

“Can I use GeoEvent Server 10.7 with my older Enterprise deployments”

- Leverage current architecture and functionality
  - Requires RESTful processing of read/writes to Feature Layers
- Does NOT support publishing of hosted services or most spatiotemporal data store operations

Things I Wish I Knew

version parity requirements

"Do I have to upgrade or wait for newer versions of the Gallery add-ons"

- Most items on the GeoEvent Gallery are compatible with GeoEvent Server 10.4 or later
  - New releases / updates are scheduled only when a component is not compatible
- It is recommended to remove and redeploy custom components during upgrading

NOTE: The release strategy for ArcGIS GeoEvent Server components delivered on the ArcGIS GeoEvent Server Gallery has been updated. Going forward, a new release will only be created when a component has a issue, is being enhanced with new capabilities, or is not compatible with new versions of ArcGIS GeoEvent Server. This strategy makes upgrades of these custom components easier since you will not have to upgrade them for every release of ArcGIS GeoEvent Server unless a new version of that connector is released. The documentation for the latest release has been updated and includes instructions for updating your configuration to align with this strategy. We apologize for any inconvenience this change in strategy causes.

Latest Release:

Release 4 - April 19, 2019: Compatible with any GeoEvent Server version 10.4 or later. (**Note:** documentation updated June 6, 2019)

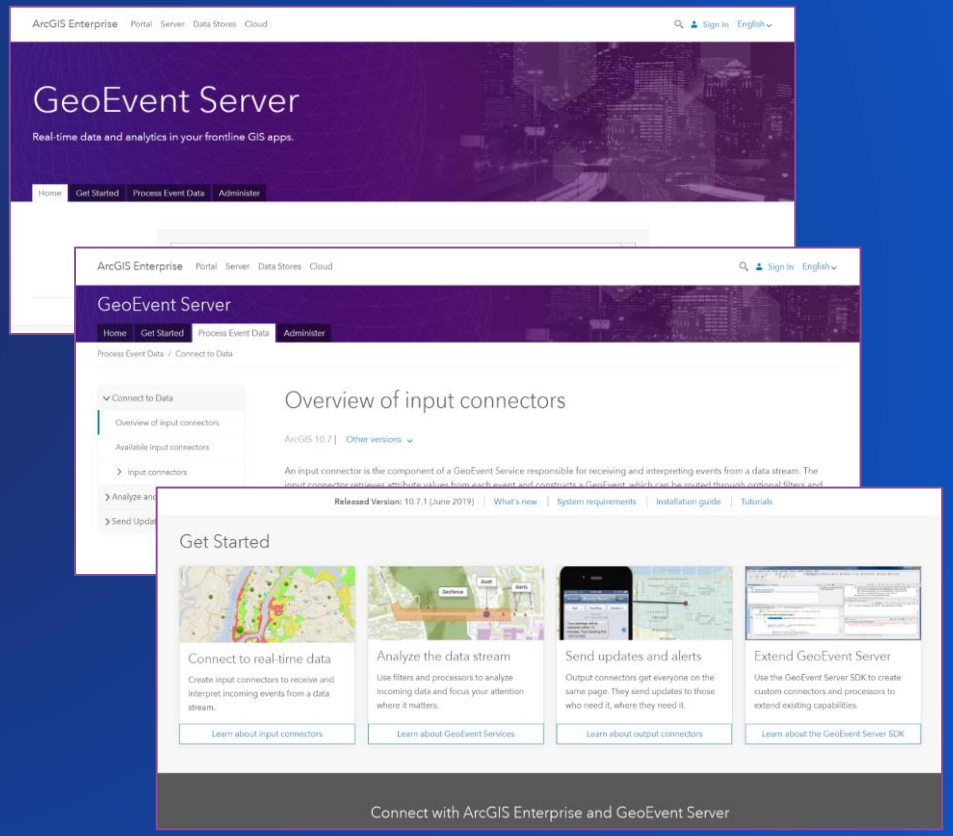
Previous Releases:

Release 3 - June 27, 2018: Compatible with any GeoEvent Server version 10.4 or later.
Release 2 - October 18, 2016: Compatible with any GeoEvent Server version 10.4 or later.
Release 1 - October 11, 2016: Compatible with any GeoEvent Server version 10.4 or later.

Things I Wish I Knew

how do I get started?

- **ArcGIS GeoEvent Server resources**
  - **http://enterprise.arcgis.com/en/geoevent**
    - **Updated Documentation**
    - **Installation Guides**
    - **System Requirements**
    - **Tutorials**
- **Blogs and discussions on the forum**
  - **http://links.esri.com/geoevent-forum**
- **Video recordings of technical workshops**
  - **http://www.esri.com/videos**

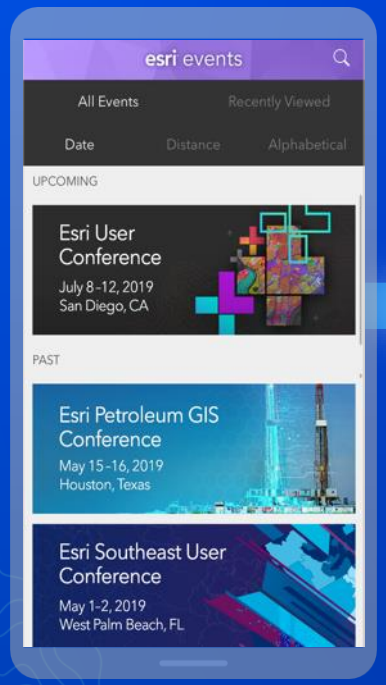


Real-Time and Big Data Technical Workshops

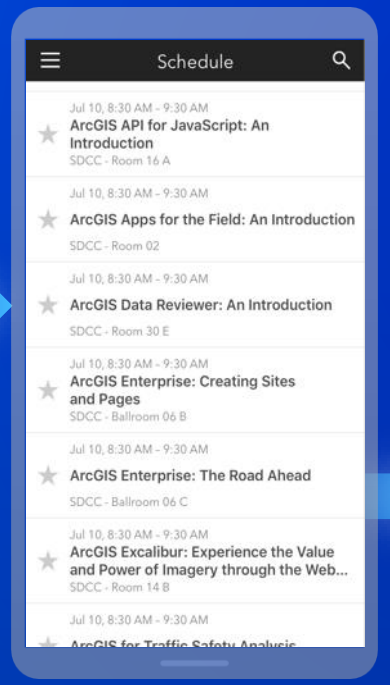
- **Wednesday**
  - 10:00 - 11:00 Real-Time & Big Data GIS: Best Practices
  - 1:00 - 2:00 ArcGIS GeoEvent Server: An Introduction — 2nd offering
  - 4:00 - 5:00 ArcGIS GeoEvent Server: Applying Real-Time Analytics — 2nd offering
  - 4:00 - 5:00 ArcGIS and the Internet of Things (IoT) — 2nd offering
- **Thursday**
  - 10:00 - 11:00 Real-Time & Big Data GIS: Best Practices — 2nd offering
  - 2:30 - 3:30 Real-Time & Big Data GIS: Road Ahead — Only Offering
  - 4:00 - 5:00 ArcGIS GeoEvent Server: Visualizing Real-Time Data — 2nd offering

Please Share Your Feedback in the App

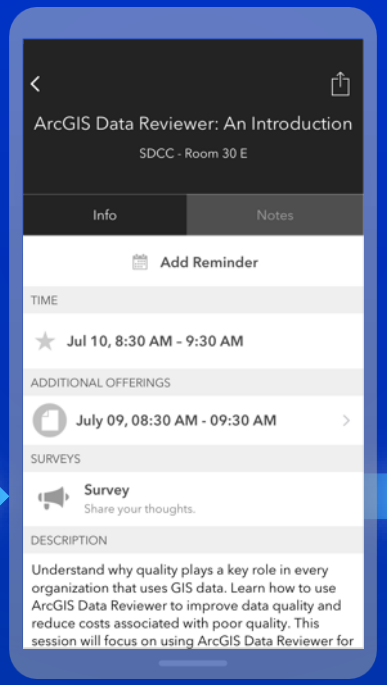
Download the Esri Events app and find your event



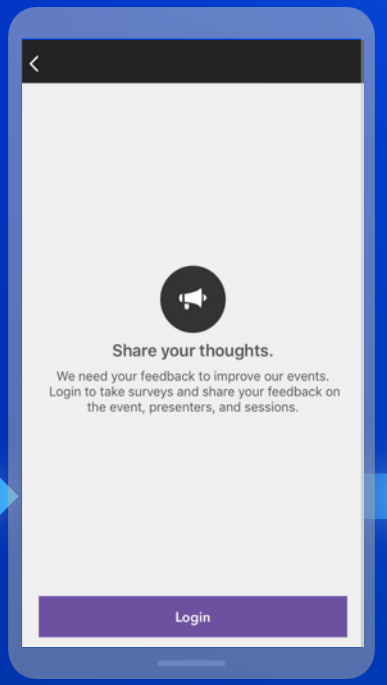
Select the session you attended



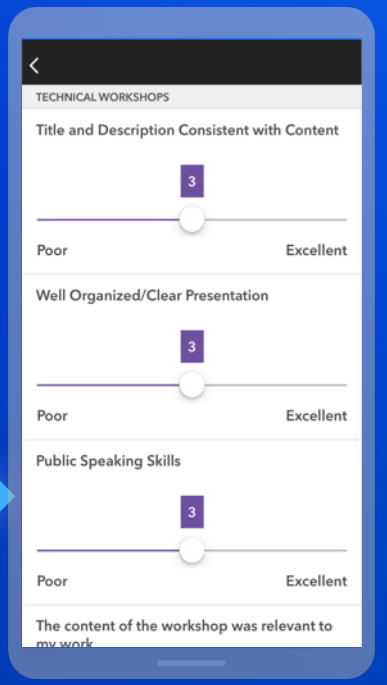
Scroll down to "Survey"



Log in to access the survey



Complete the survey and select "Submit"



Questions
Feedback

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

RJ Sunderman
ArcGIS GeoEvent Server
Product Engineer, Esri
rsunderman@esri.com

SEE
WHAT
OTHERS
CAN'T


esri
THE
SCIENCE
OF
WHERE