Big Data: ArcGIS Spatiotemporal Big Data Store

Mark Bramer

Esri Professional Services

mbramer@esri.com

2016 ESRI
PETROLEUM
GIS CONFERENCE

Agenda

1. Evolution of Real-Time GIS
2. Observation data
3. Visualizing observation data
4. ArcGIS Data Store
5. Storing observation data
6. A deeper dive
7. Capacity planning

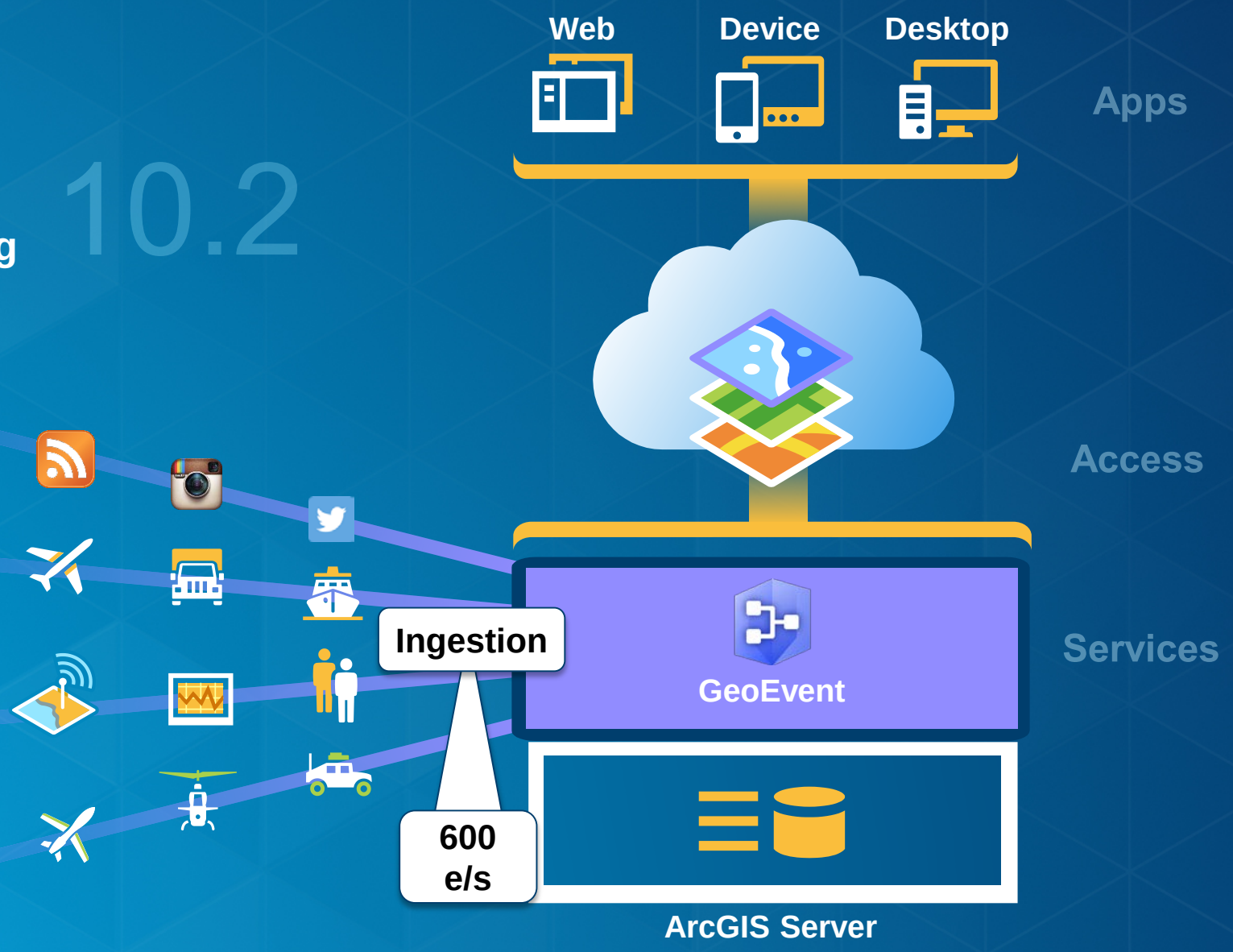
1. Evolution of Real-Time GIS

Real-Time GIS

ArcGIS 10.2

10.2

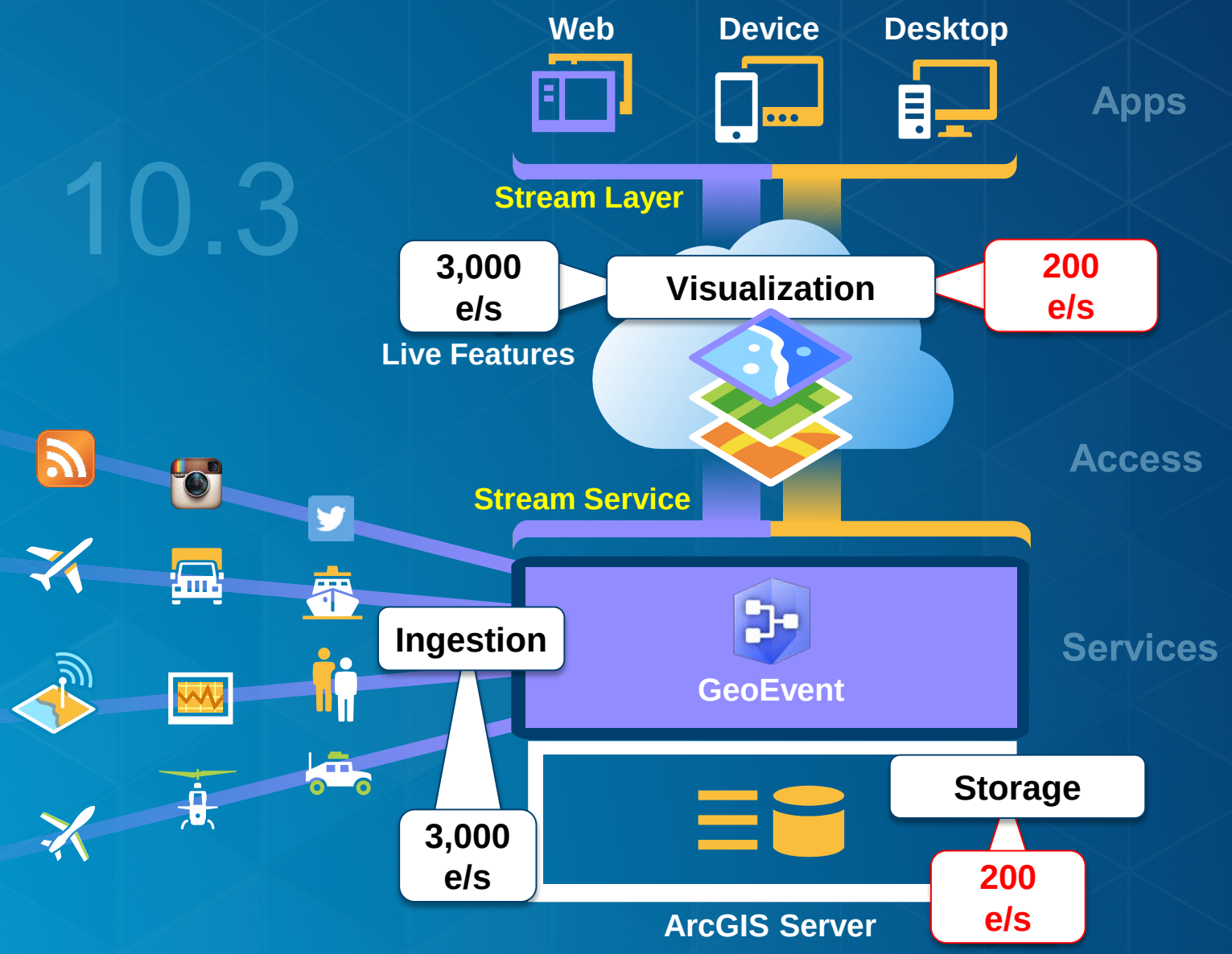
- Provide the ability to ingest real-time streaming data into ArcGIS.
- Perform real-time analytics on event data as it is received.
- Send updates and alerts to those who need it, where they need it.



Real-Time GIS

ArcGIS 10.3

- Can ingest real-time high velocity data into ArcGIS but low volume storage.
- Can visualize high velocity & volume data in raw feature form live, but not historically.
- Visualization can't scale beyond known constraints.

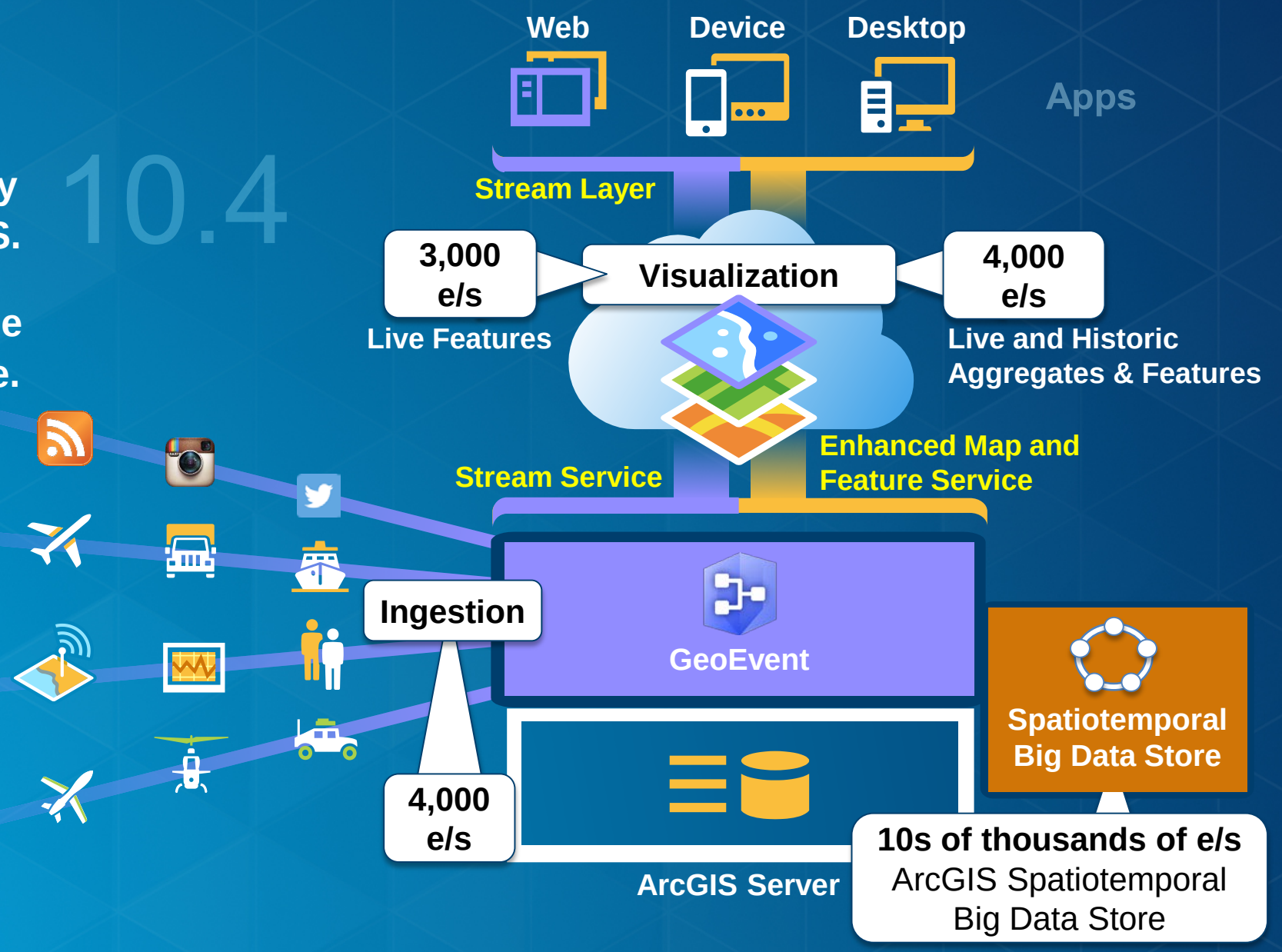


Real-Time GIS

ArcGIS 10.4

10.4

- Can ingest higher velocity real-time data into ArcGIS.
- Observations CAN now be stored in a Big Data Store.
- Can visualize high velocity and volume data
  - as an AGGREGATION,
  - as discrete FEATURES,
  - live & HISTORICALLY.
- Visualization CAN scale.



2. Observation Data

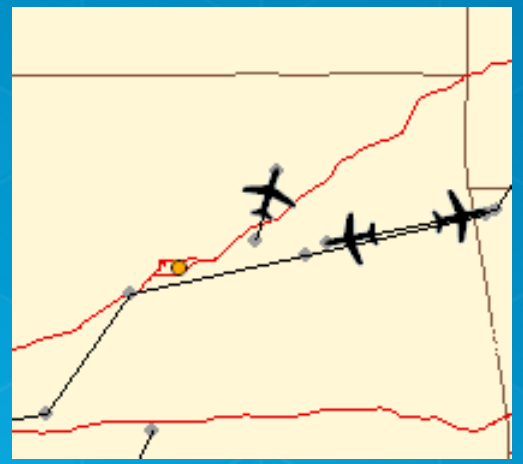
Observation Data

Types

- Types of observation data include:

Moving

something that moves



- Planes
- Vehicles
- Animals
- Satellites
- Storms

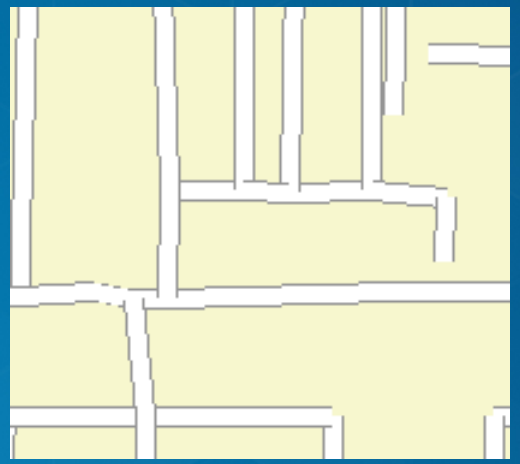
Stationary

stands still but attributes change

- Water Gauges
- Weather Stations
- Traffic Sensors
- Air Quality

Discrete

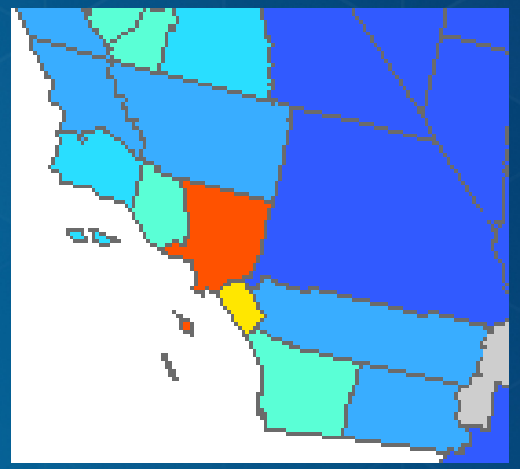
something that "just happens"



- Crimes
- Lightning
- Accidents

Change

change or growth



- Population
- Distribution
- Fire Perimeters

Observation Data

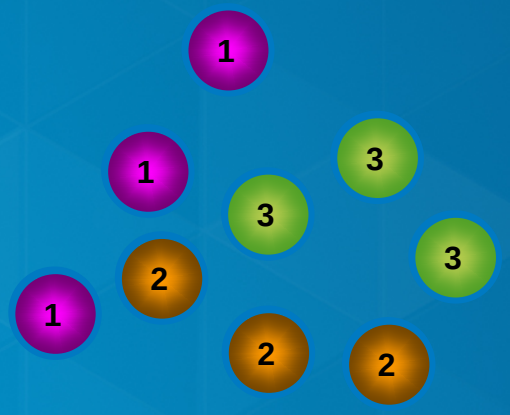
Defined

- **An observation is a recording of a feature's attribute values and location at a specific moment in time.**
  - **Observations can be replayed over space & time.**
  - **Moving observations are identifiable by a unique attribute, known as a TRACK_ID.**

Space & Time

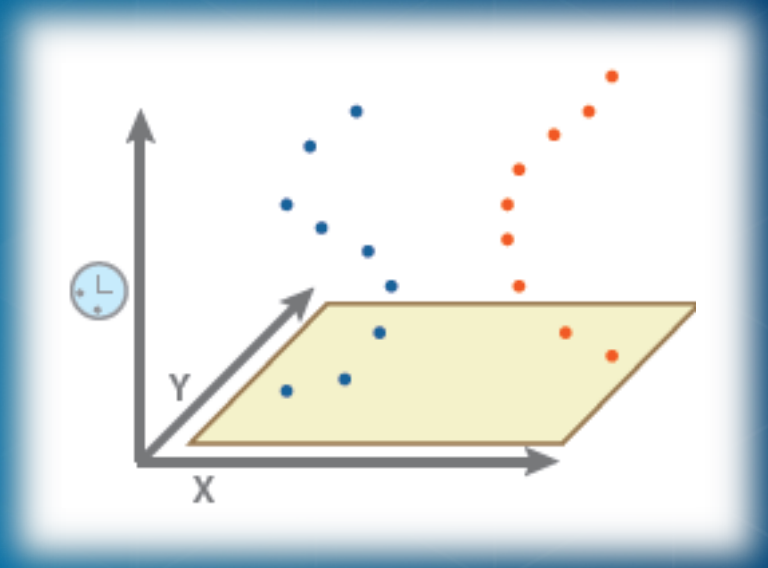
Observations can be shown at a specific time

Tracks



Moving observations can be identified by a unique attribute

Observations are like cells in a filmstrip



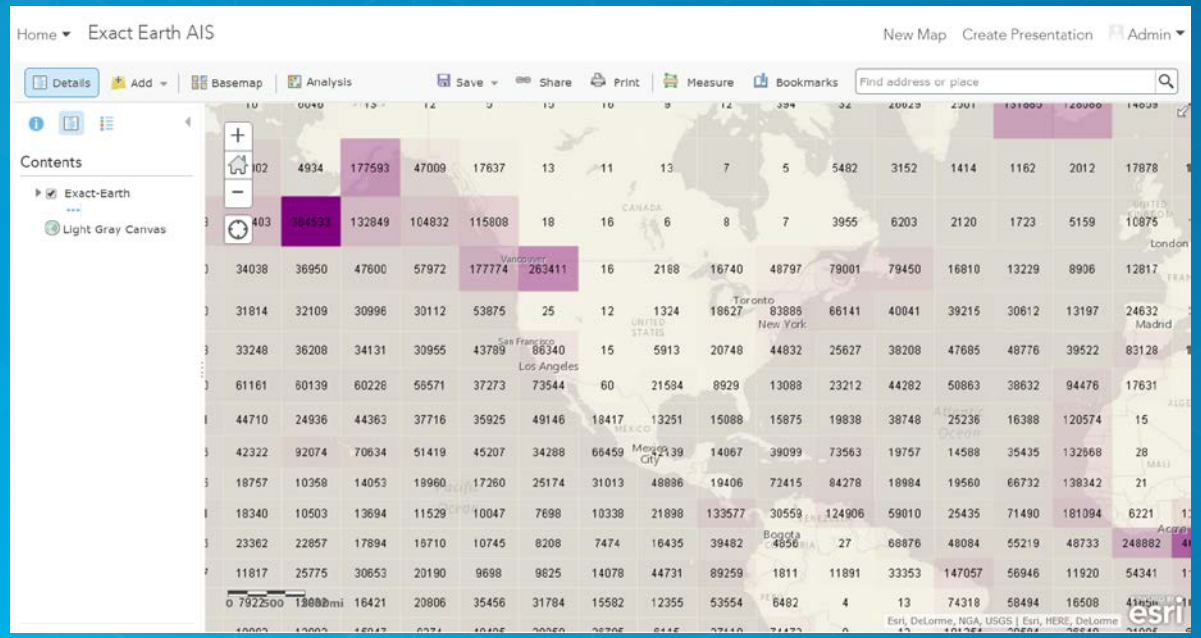
Moving observation illustration for two tracks over space (x, y) and time (t)

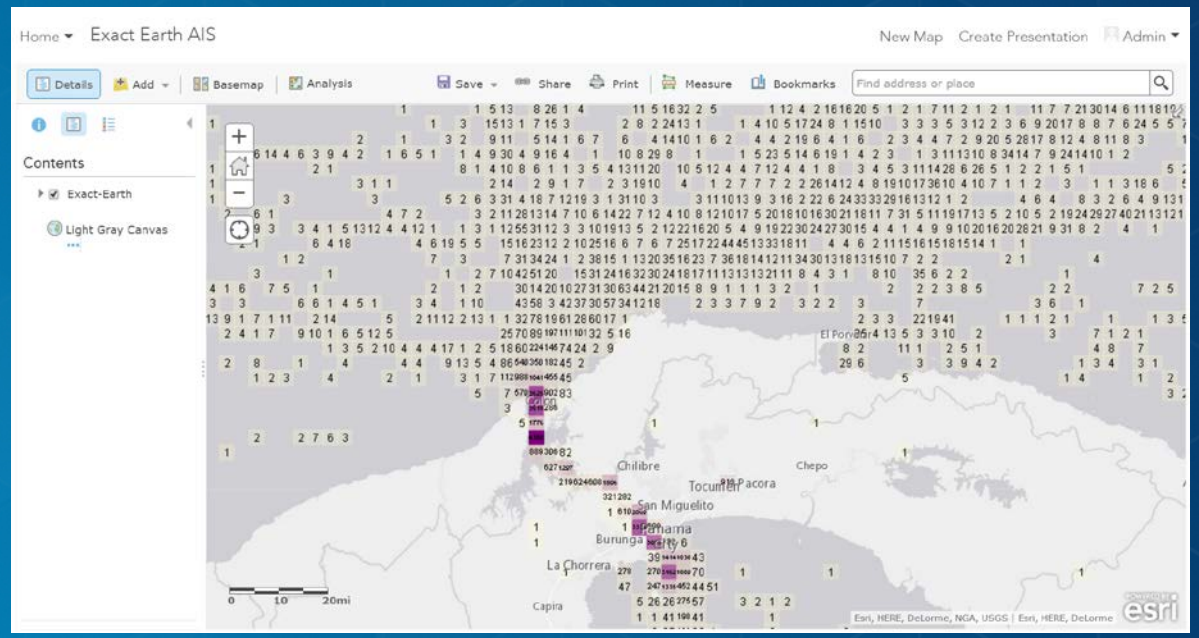
3. Visualizing Observation Data

Visualizing Observation Data

ArcGIS for Server: Map Service

- **High volume observation data can be visualized as an aggregation**
  - **An on-the-fly aggregation (representation) of the data is prepared by the map service**
  - **Clients querying for high volumes of data are given a rendered representation of the data**
  - **Aggregations are calculated at various levels of detail and are specific to each user session.**

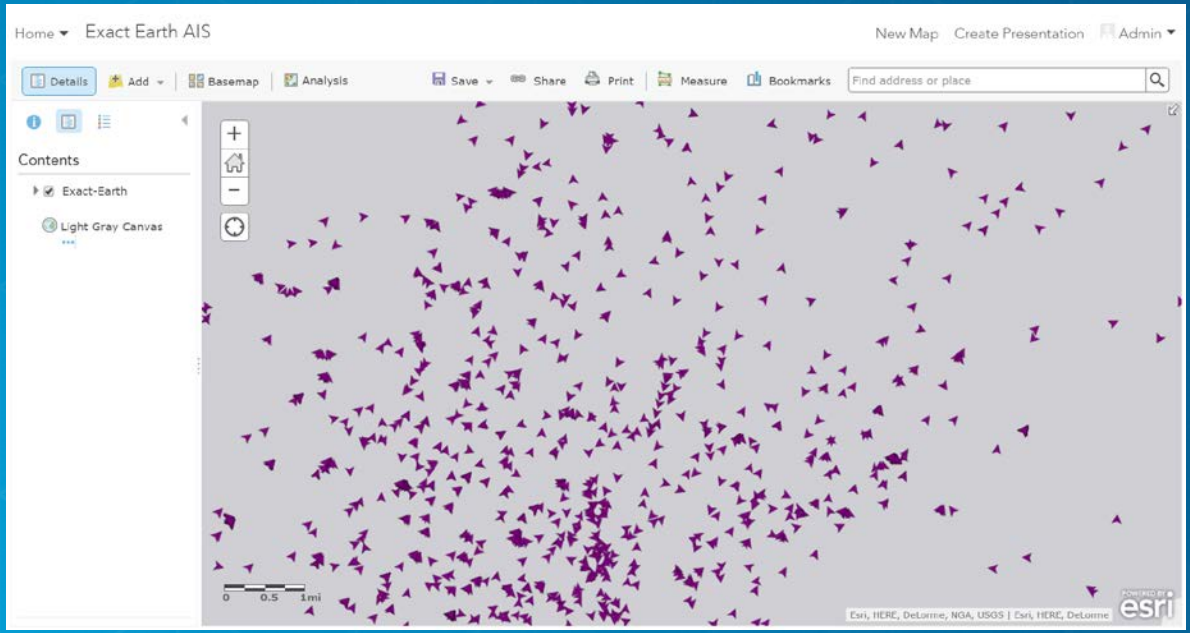




Visualizing Observation Data

ArcGIS for Server: Map Service

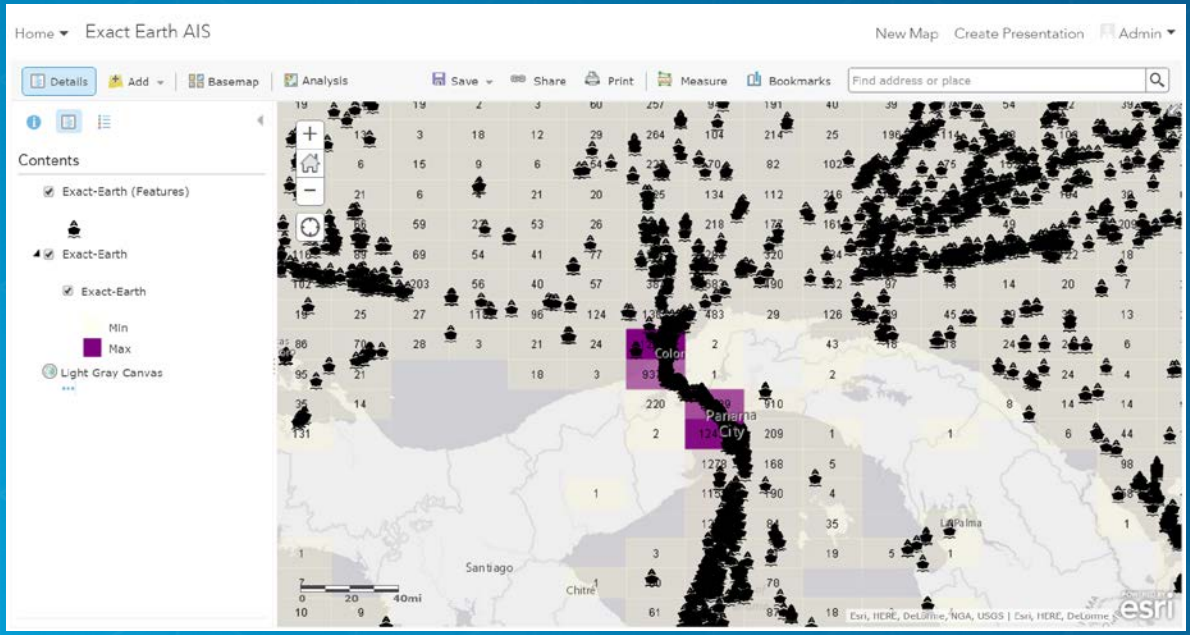
- **The map service's aggregation can be configured to display discrete features**
  - **Big Data visualization is scalable based on the density of your data**
  - **Specify a *Feature Threshold* and filter to display a reasonable number of features at any map scale**



Visualizing Observation Data

ArcGIS for Server: Feature Service

- **High volume observation data can be visualized as discrete features**
  - **Data queried as features from a big data store can be symbolized by a client**
  - **Combine a map service and feature service for traditional scale dependent rendering**



Visualizing Observation Data

ArcGIS for Server: Feature Service

- Efficiently query your observations with fast access
  - Query by any combination of id, time, space, and/or attribute value

- 1 Month worth of AIS shipping vessel data
  - 28,243,036 observations

- Query one vessel by ID (MMSI = ‘440411000’)
  - Within the time extent
    - From Jan 9, 2012 4:07:48 PM PDT
    - To Jan 23, 2012 7:27:08 PM PDT
  - Intersecting a spatial envelope
    - xmin: -117, ymin: 28 and xmax: -75, ymax=45, srid=4326

Visualizing Observation Data

ArcGIS for Server: Feature Service

10.4

Query: Exact-Earth (ID: 0)

Where:	MMSI = '440411000'
Text:	
Object IDs:	
Time:	1326067668000,1327375628000
Input Geometry:	{"xmin": -117, "ymin": 28, "xmax": -75, "ymax": 45, "spatialReference": {"wkid": 4326} }
Geometry Type:	Envelope
Input Spatial Reference:	
Spatial Relationship:	Intersects
Relation:	
Out Fields:	*
Return Geometry:	◉ True ○ False
Format:	HTML

Query (GET) Query (POST)

60 out of **28,243,036** in **83 ms**!

 # records: 60

4. ArcGIS Data Store

4a. ArcGIS Data Store

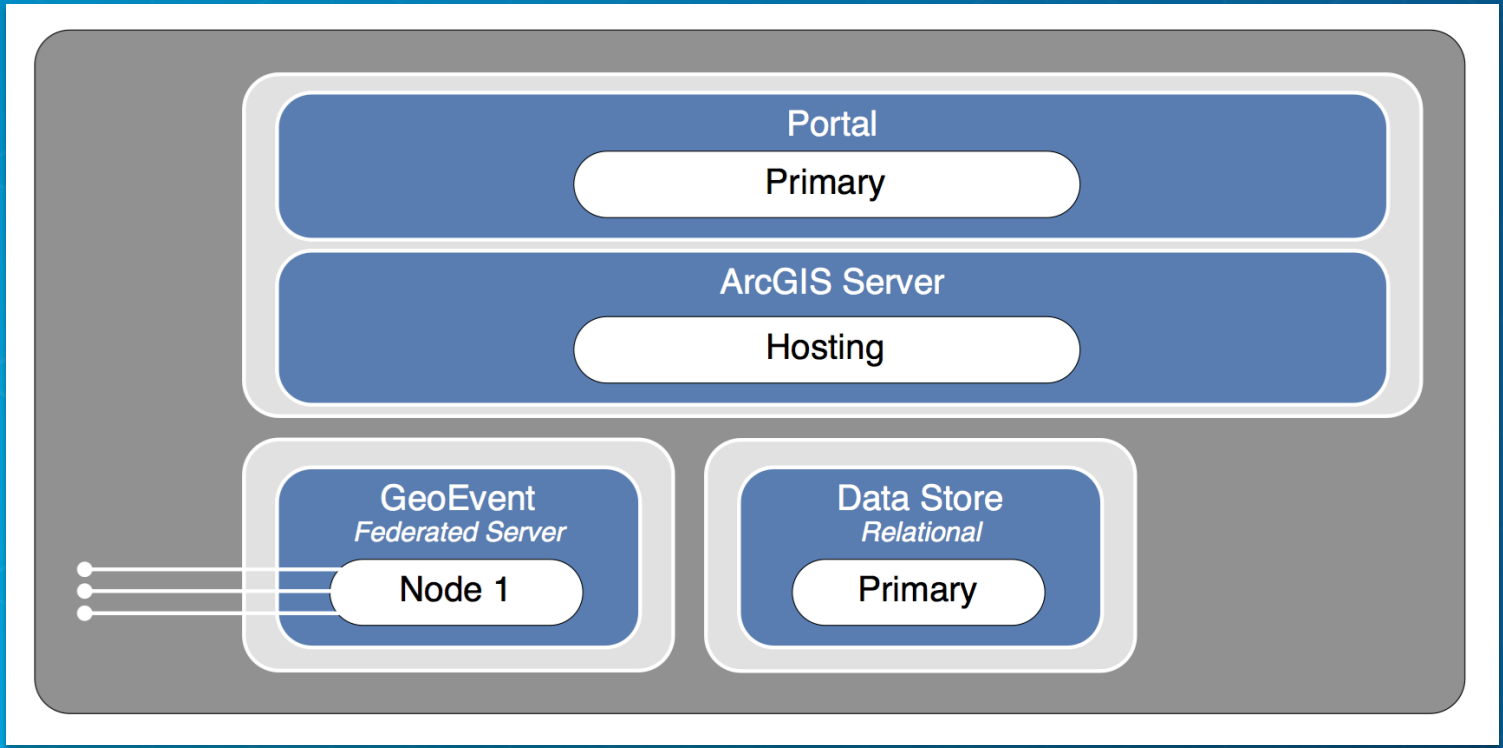
Relational Data Store

ArcGIS Data Store

Relational Data Store

10.3

- The ArcGIS 10.3 Data Store enables the storage of items via a Managed Database.
- The Relational Data Store enables the storage of
  - traditional Feature Services, Map Services, and many other items.

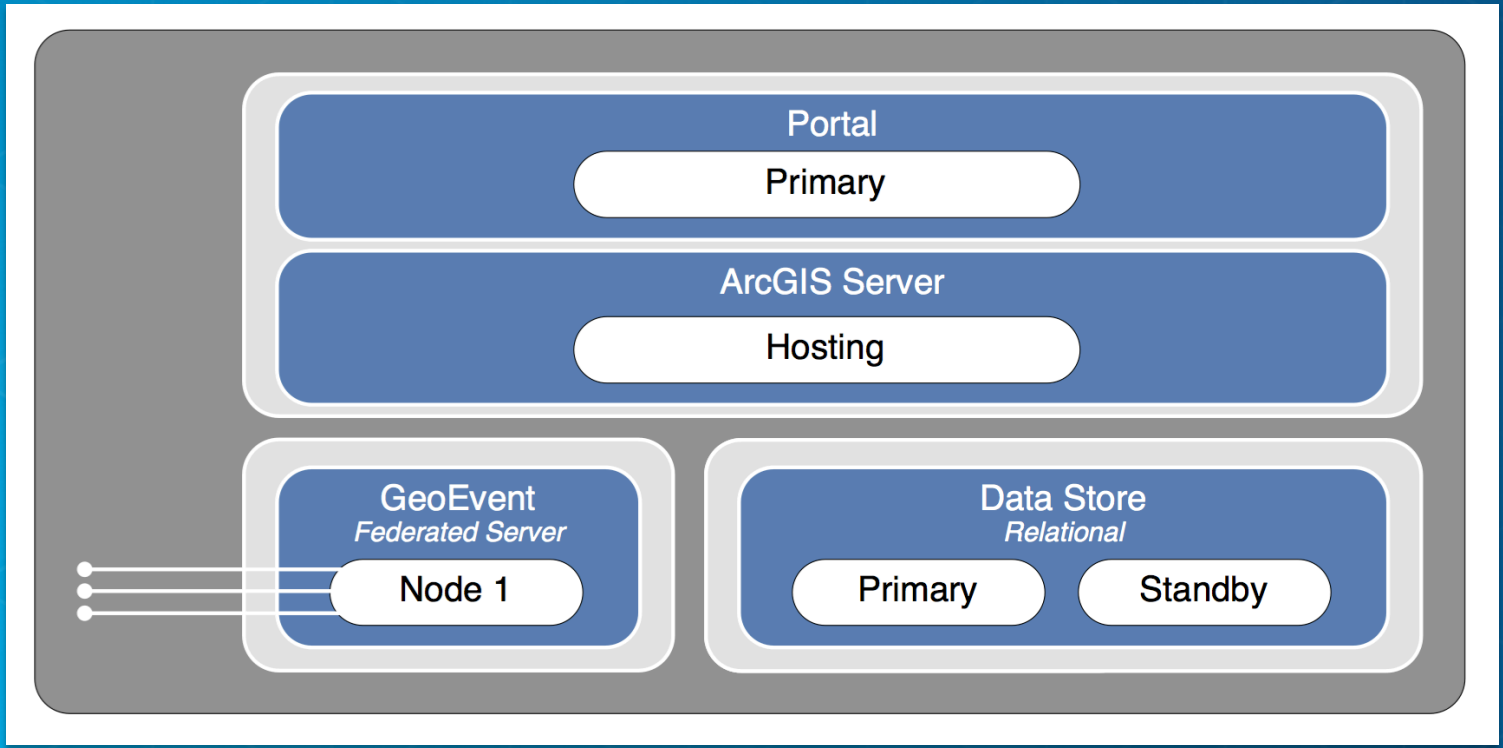


10.3

ArcGIS Data Store

Relational Data Store

- The ArcGIS 10.3 Data Store enables the storage of items via a Managed Database.
- The Relational Data Store enables the storage of
  - traditional Feature Services, Map Services, and many other items.
- You can enable high availability on the Relational Data Store by adding a standby.



4b. ArcGIS Data Store

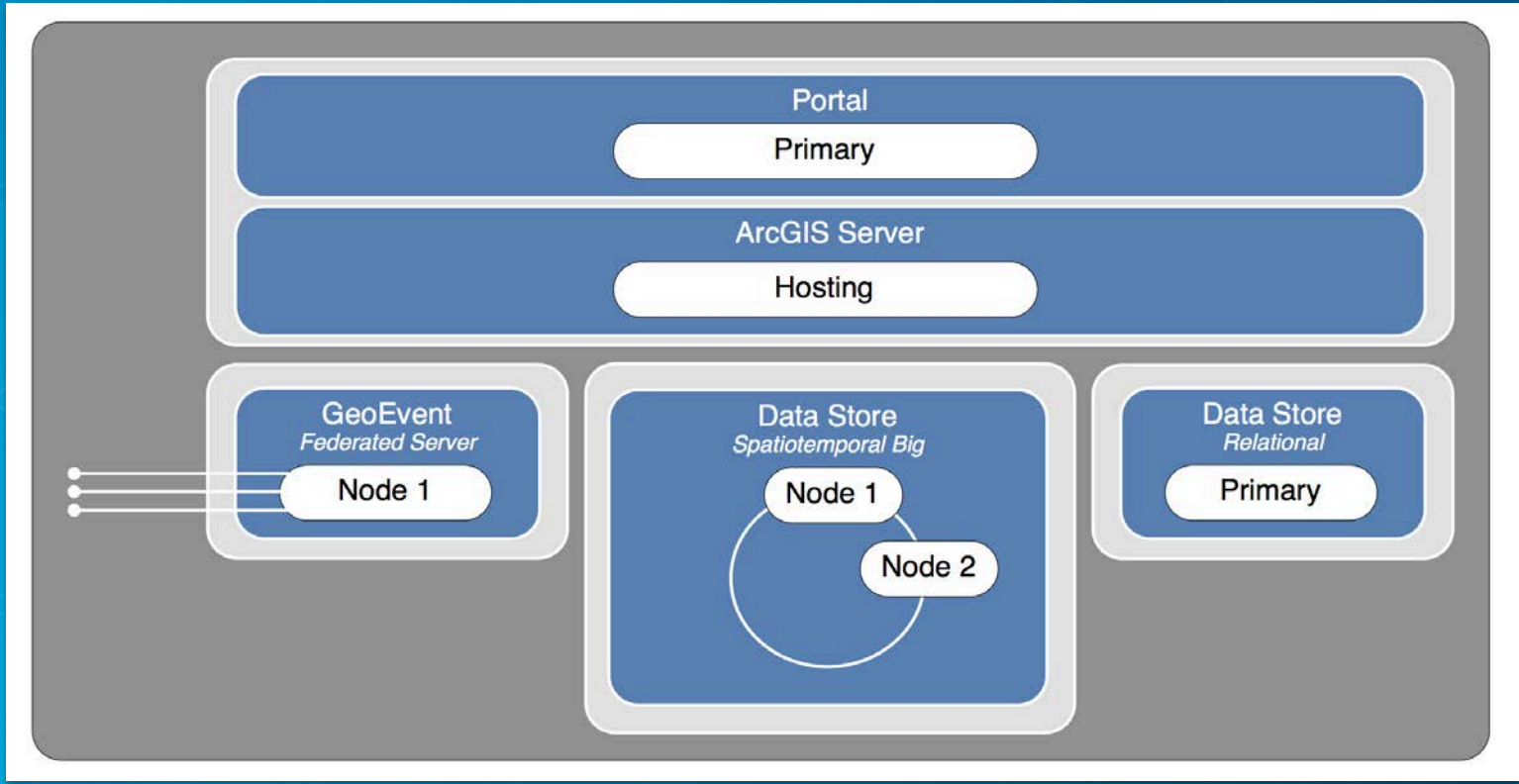
Spatiotemporal Big Data Store

ArcGIS Data Store

Spatiotemporal Big Data Store

10.4

- **The ArcGIS 10.4 Data Store now works with high velocity & volume observation data.**
- **The Big Data Store enables storage and archival of spatiotemporal observation data.**
  - **it can sustain tens of thousands of writes per second.**

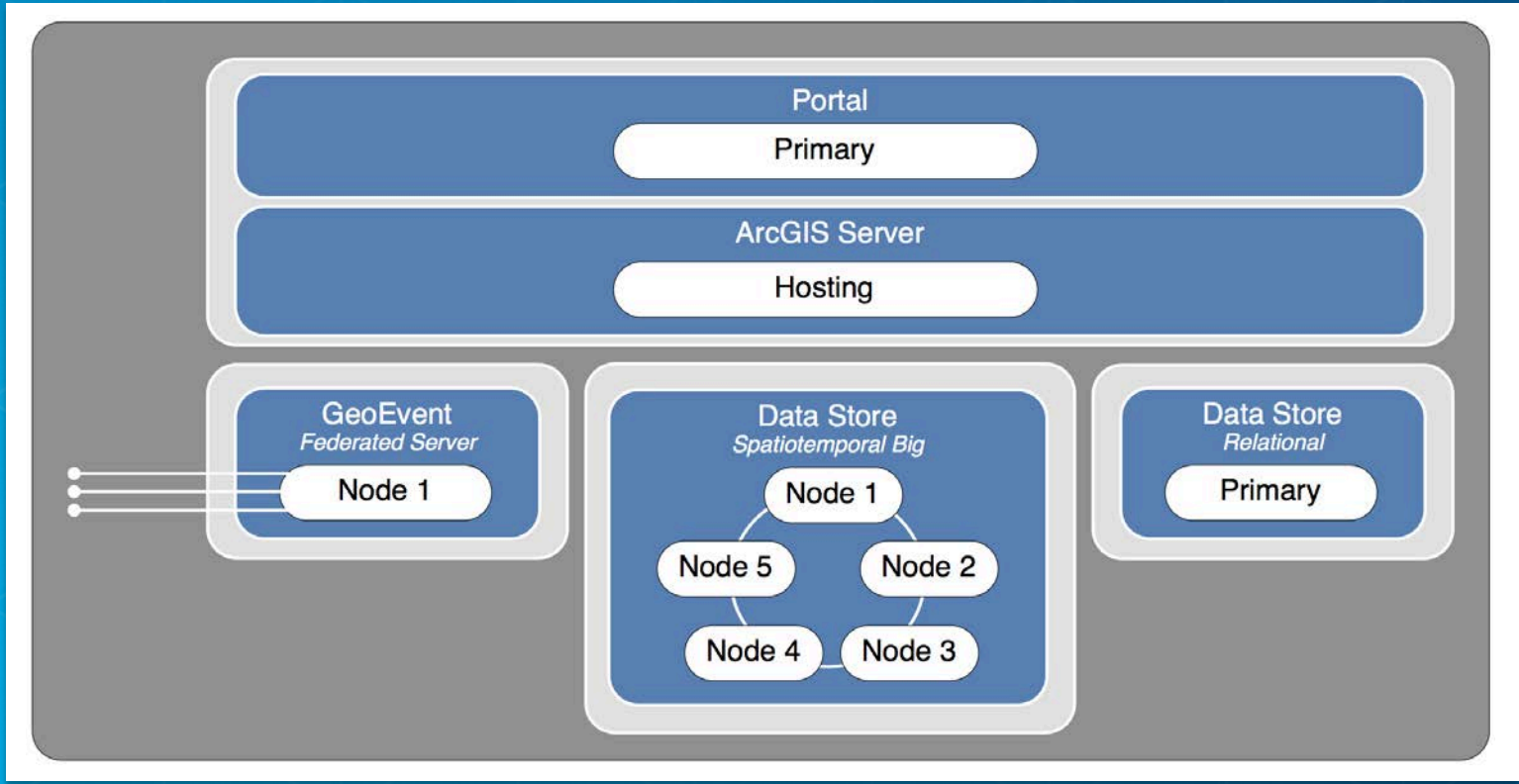


10.4

ArcGIS Data Store

Spatiotemporal Big Data Store

- **The ArcGIS 10.4 Data Store now works with high velocity & volume observation data.**
- **The Big Data Store enables storage and archival of spatiotemporal observation data.**
  - **it can sustain tens of thousands of writes per second.**
- **You can enable higher volume storage on the Big Data Store by adding additional nodes.**

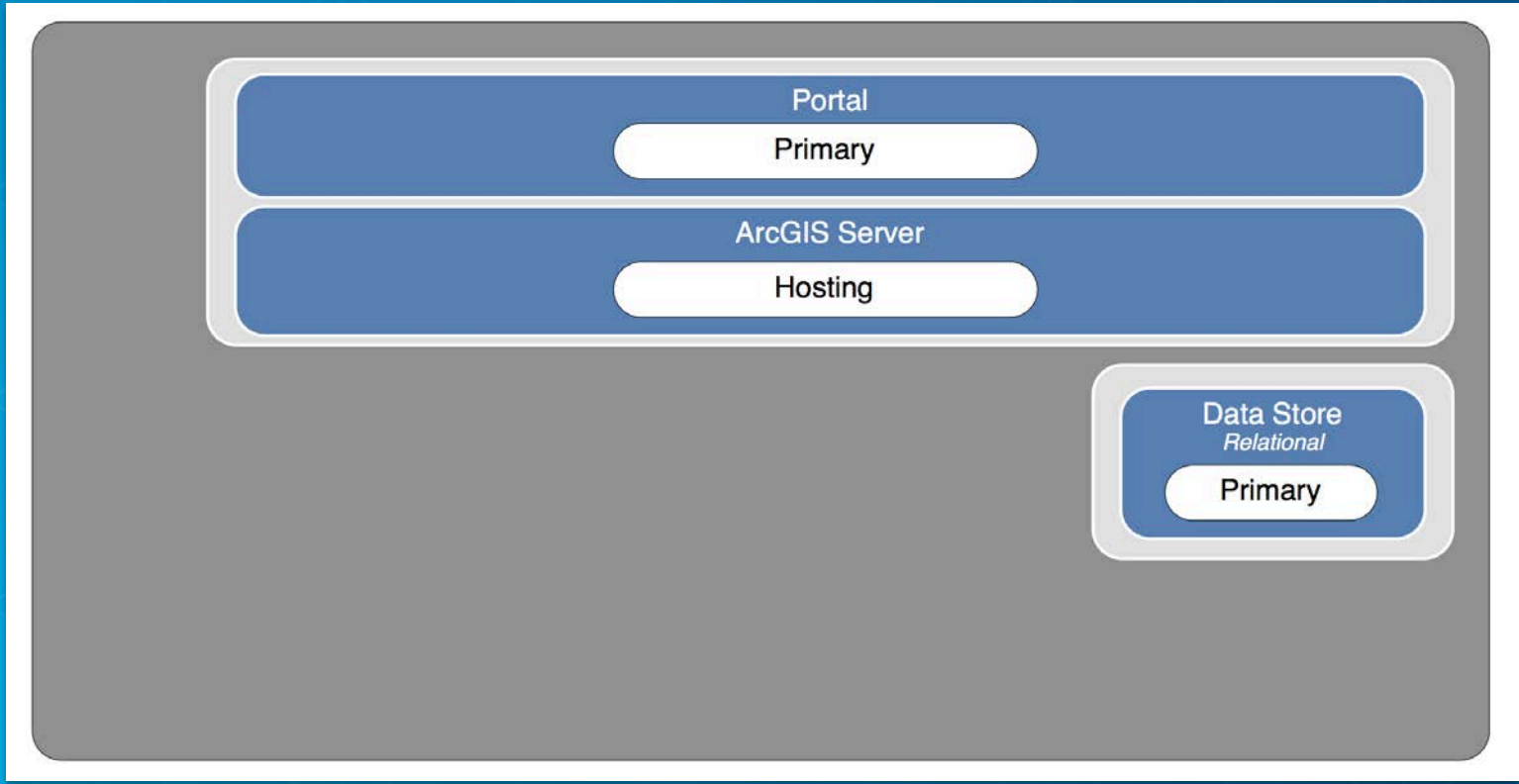


10.4

Installation Experience

Spatiotemporal Big Data Store Pre-requisites

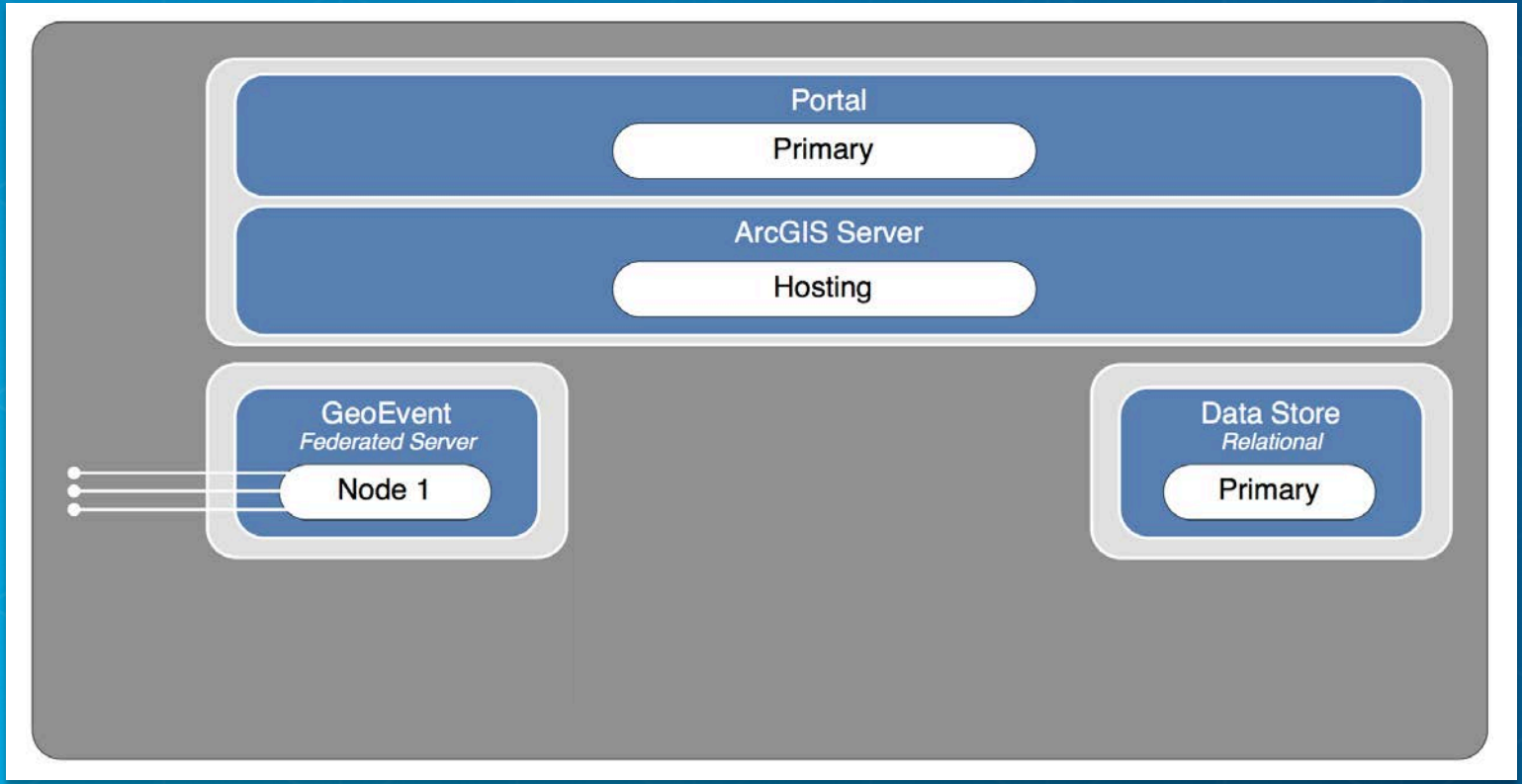
- **Portal for ArcGIS has been installed and configured.**
- **ArcGIS Data Store has been installed and registered with an ArcGIS Server as a Relational Data Store.**
- **An ArcGIS Server has been registered with Portal as a “Hosting” server.**



Installation Experience

Spatiotemporal Big Data Store Pre-requisites (continued)

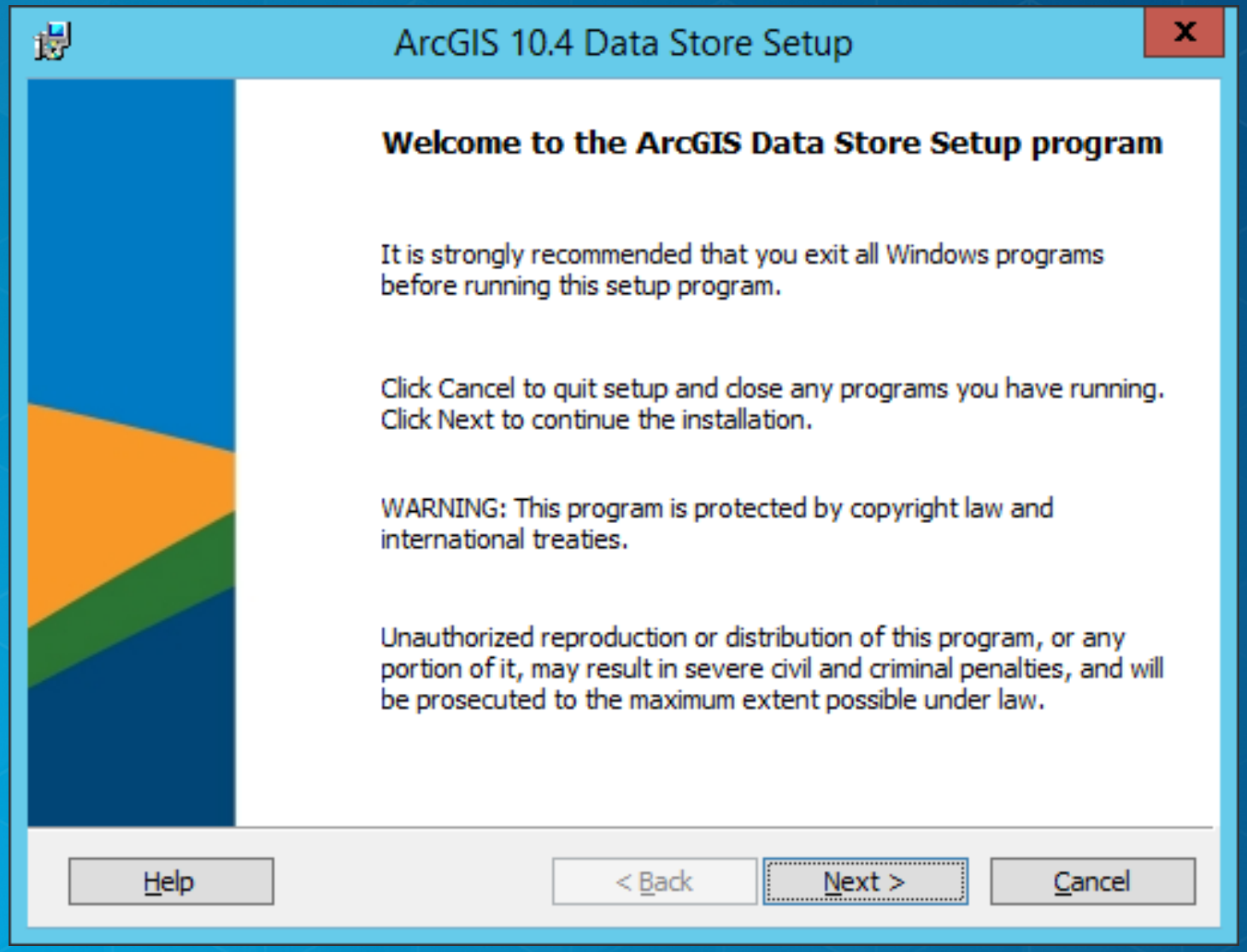
- **An ArcGIS Server with GeoEvent Extension for Server has been registered with Portal as a "Federated" server.**
- **Configure SSL security certificates.**



Installation Experience

ArcGIS Data Store

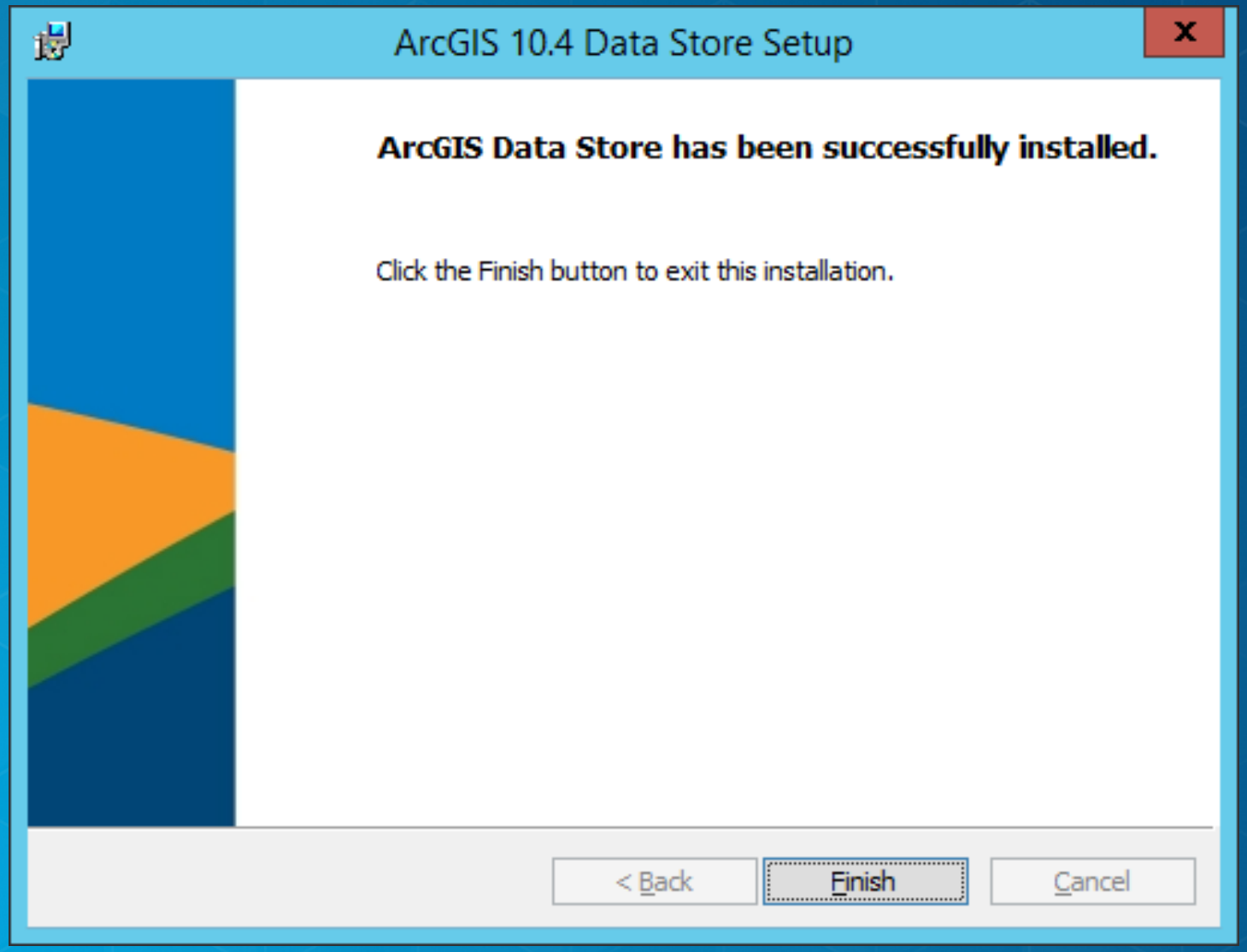
10.4



Installation Experience

ArcGIS Data Store

10.4



Installation Experience

Enabling the Spatiotemporal Big Data Store

10.4

```
cd %AGSDATASTORE%\tools

configuredatastore.bat https://webgis-server:6443/arcgis/admin your_username
your_password C:\arcgisdatastore --stores spatiotemporal
```

10.4

Installation Experience

Enabling the Spatiotemporal Big Data Store

Administrator: Command Prompt - configuredatastore.bat https://ps000257.esr...

```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\        >cd\

C:\>cd "Program Files\ArcGIS\DataStore\tools"

C:\Program Files\ArcGIS\DataStore\tools>configuredatastore.bat https://ps000257.
esri.com:6443/arcgis/admin admin admin C:\arcgisdatastore --stores spatiotempora
l
```

Installation Experience

Enabling the Spatiotemporal Big Data Store

Administrator: Command Prompt

```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\        >cd\

C:\>cd "Program Files\ArcGIS\DataStore\tools"

C:\Program Files\ArcGIS\DataStore\tools>configuredatastore.bat https://ps000257.
esri.com:6443/arcgis/admin admin admin C:\arcgisdatastore --stores spatiotempora
l

Configuring data store(s). It may take a few minutes. Please wait...


Operation completed successfully


C:\Program Files\ArcGIS\DataStore\tools>_
```

Verifying Installation

Admin API: nosqlDatabases

10.4

Data Item (/nosqlDat

https://localhost:6443/arcgis/admin/data/items/nosqlDatabases/AGSDataStore_bigdata_ds_ro7qu6az

ArcGIS Server Administrator Directory Logged in: admin [Administrator] | Signout

Home > data > items > nosqlDatabases > AGSDataStore_bigdata_ds_ro7qu6az API Reference

Data Item (/nosqlDatabases/AGSDataStore_bigdata_ds_ro7qu6az)

Data Item Properties

Item Path: /nosqlDatabases/AGSDataStore_bigdata_ds_ro7qu6az

Item Type: nosql

ID: 53beadeb-8ae1-444f-8993-5b8d98f2036a

Client Path: N/A

Resources: machines

Supported Operations: edit

Supported Interfaces: REST

Verifying Installation

Admin API: nosqlDatabases

10.4

https://localhost:644

https://localhost:6443/arcgis/admin/data/items/nosqlDatabases/AGSDataStore_bigdata_ds_ro7qu6az?f=pjson

Toggle Collapsed | use cfdump format:

Query: - JSONQuery

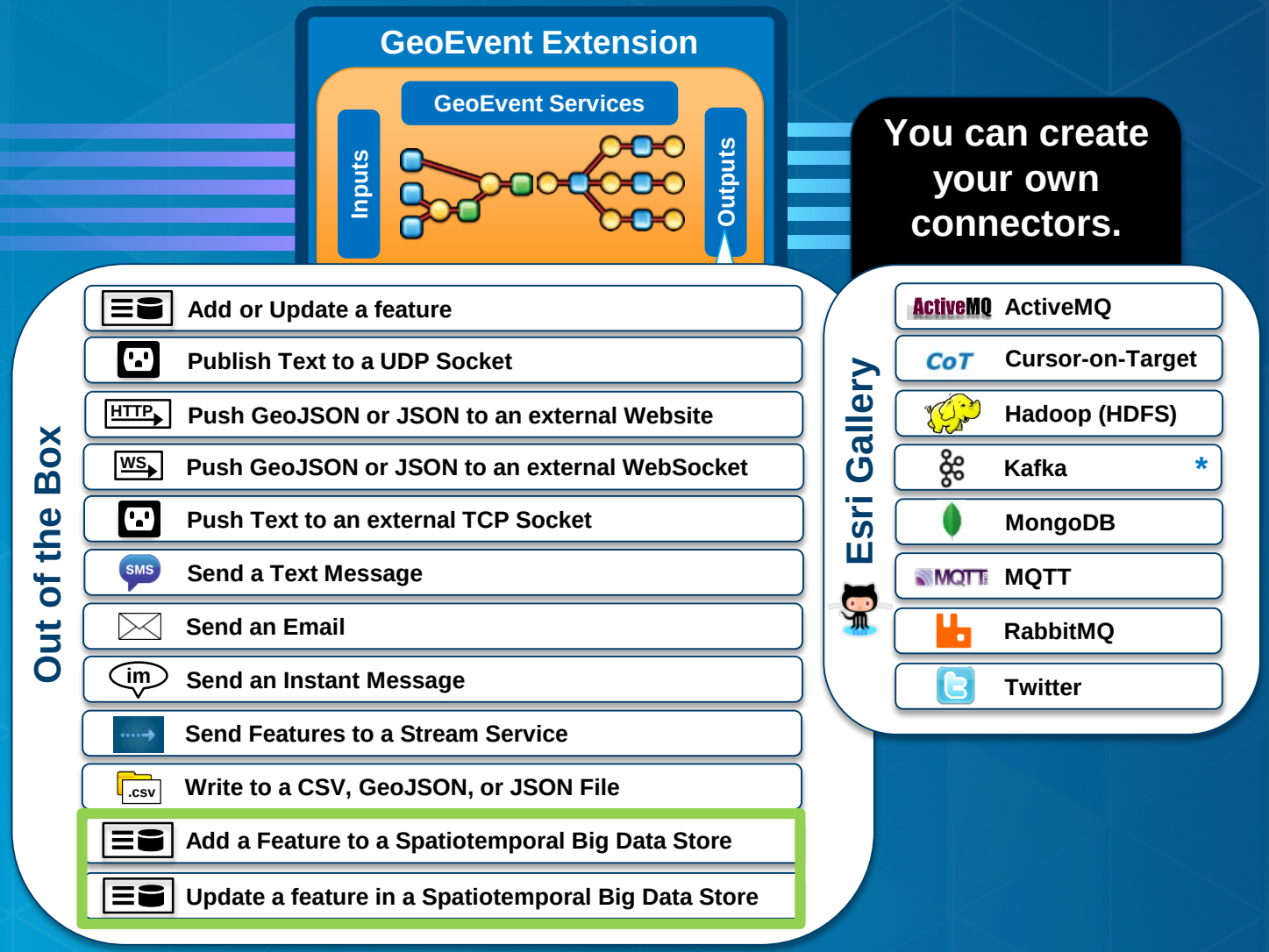

```
{
    path: "/nosqlDatabases/AGSDataStore_bigdata_ds_ro7qu6az",
    type: "nosql",
    id: "53beadeb-8ae1-444f-8993-5b8d98f2036a",
    provider: "ArcGIS Data Store",
  - info: {
        clusterName: "ds_ro7qu6az",
        dsFeature: "spatioTemporal",
        storeSig: "vpNXSaS7lqzBF8yCCCovpA== ",
        isManaged: true,
        datastoreName: "ds_ro7qu6az",
        datastoreAlias: "agsbigdatastore-1",
      - users: [
          - {
                username: "els_8lvgrzj",
                password: "                                ",
                role: "admin"
            }
        ],
      - machines: [
          - {
                role: "BIGDATA_ONLY",
                name: "        .ESRI.COM",
                httpPort: 9220,
                transportPort: 9320,
                dbPort: 9320,
                adminURL: "https://        .ESRI.COM:2443/arcgis/datastoreadmin/",
                platform: "Windows",
                isMasterNode: true,
                isDataNode: true
            }
        ]
    }
}
```

5. Storing Observation Data

Storing Observation Data

ArcGIS GeoEvent Extension for Server: output connectors

10.4



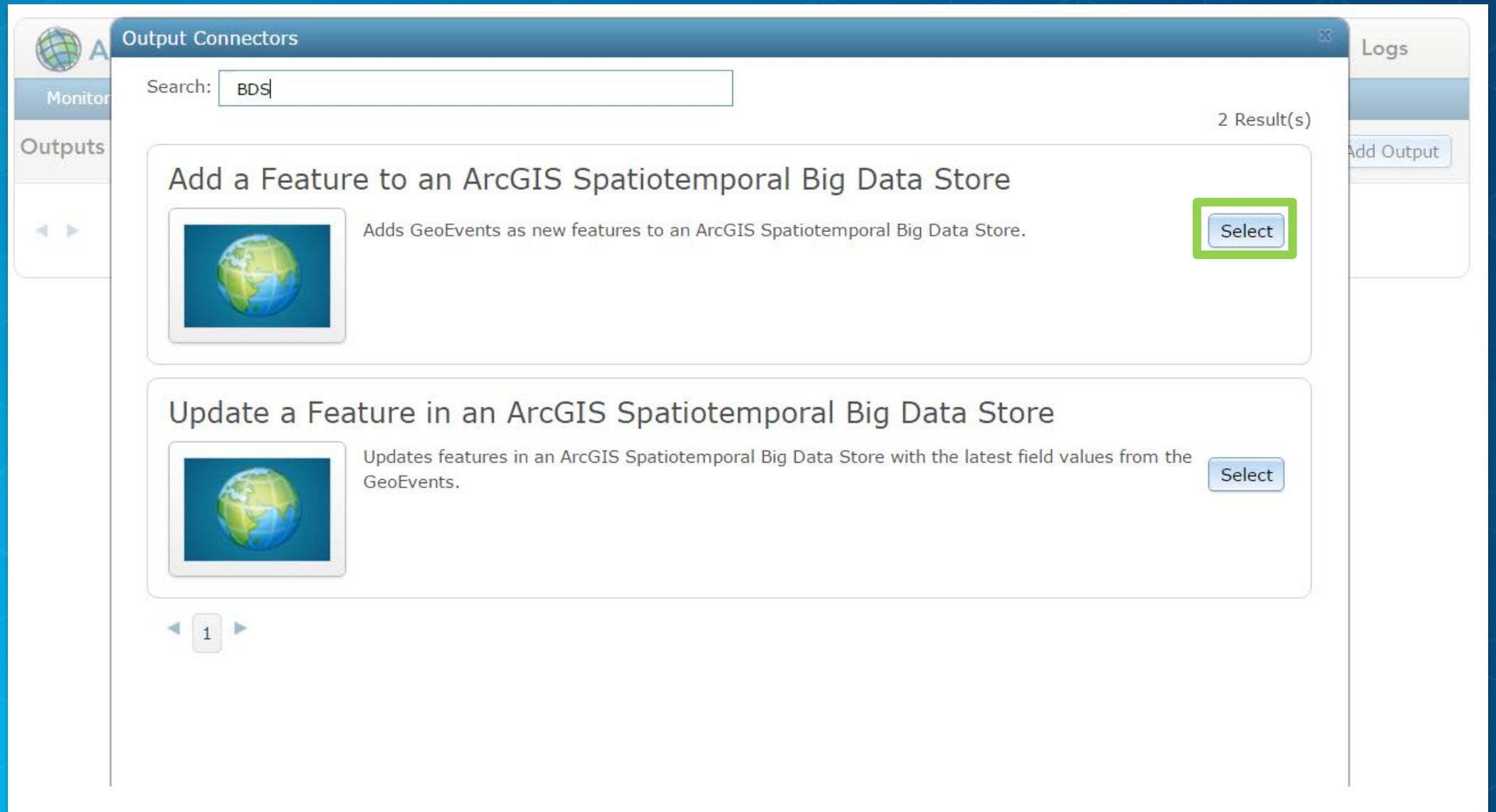
5a. Storing Observation Data

Add a Feature

Storing Observation Data

10.4

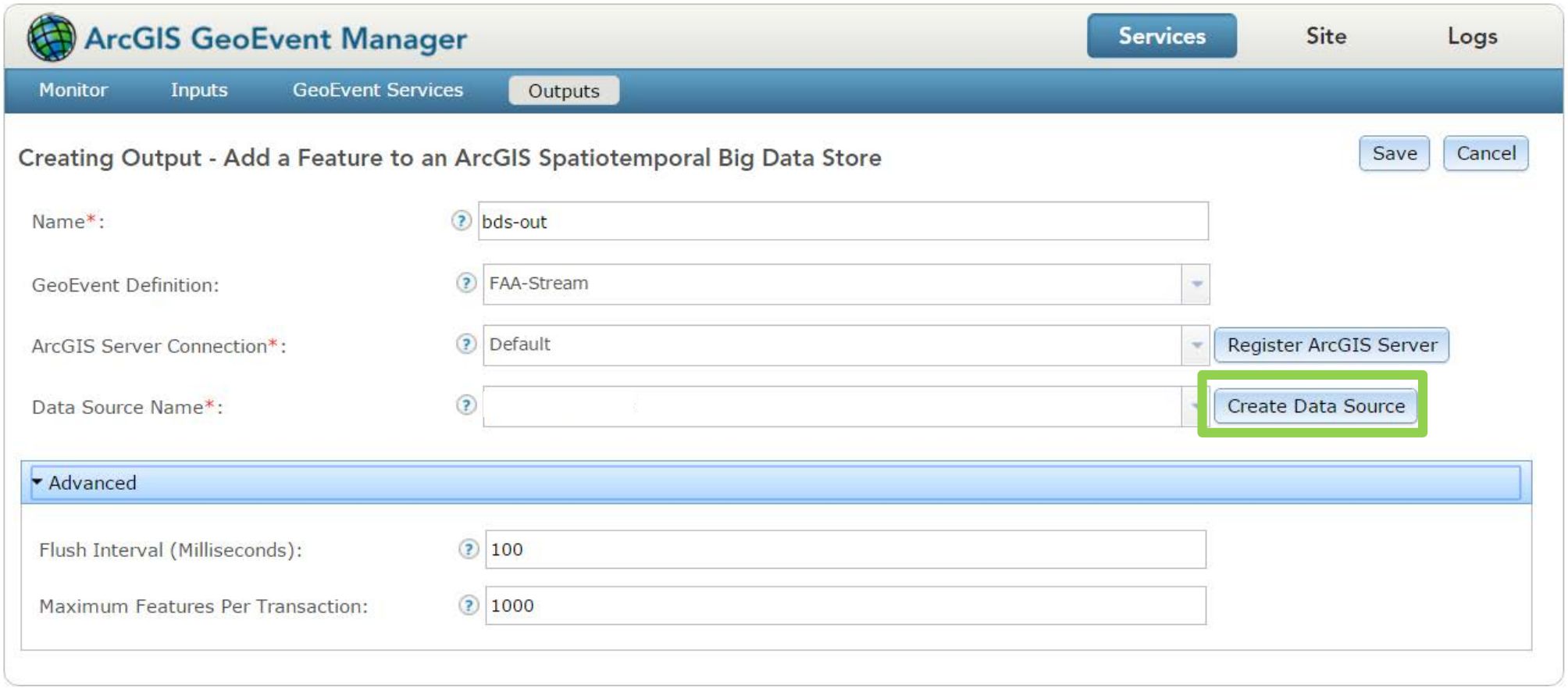
Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

10.4

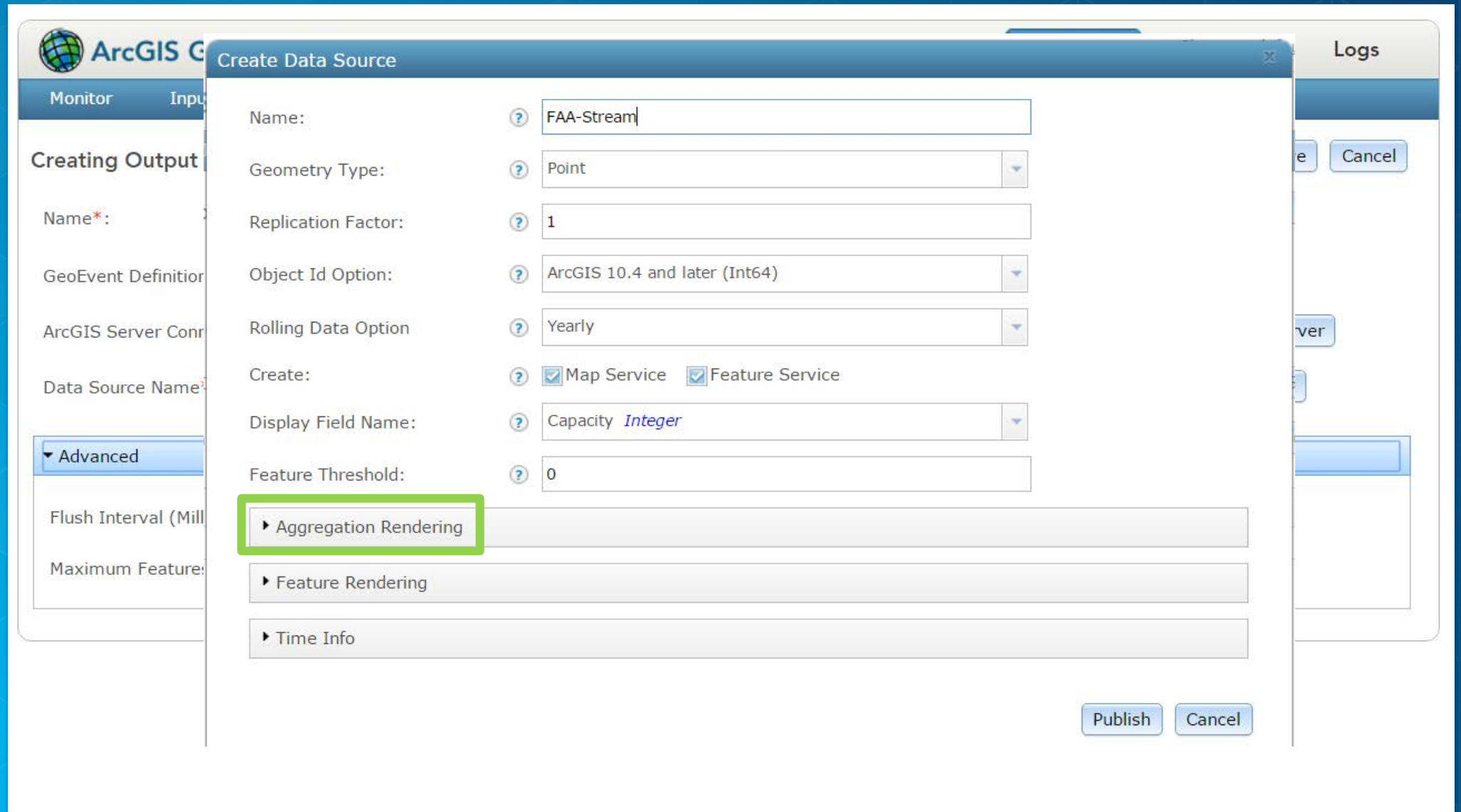
Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

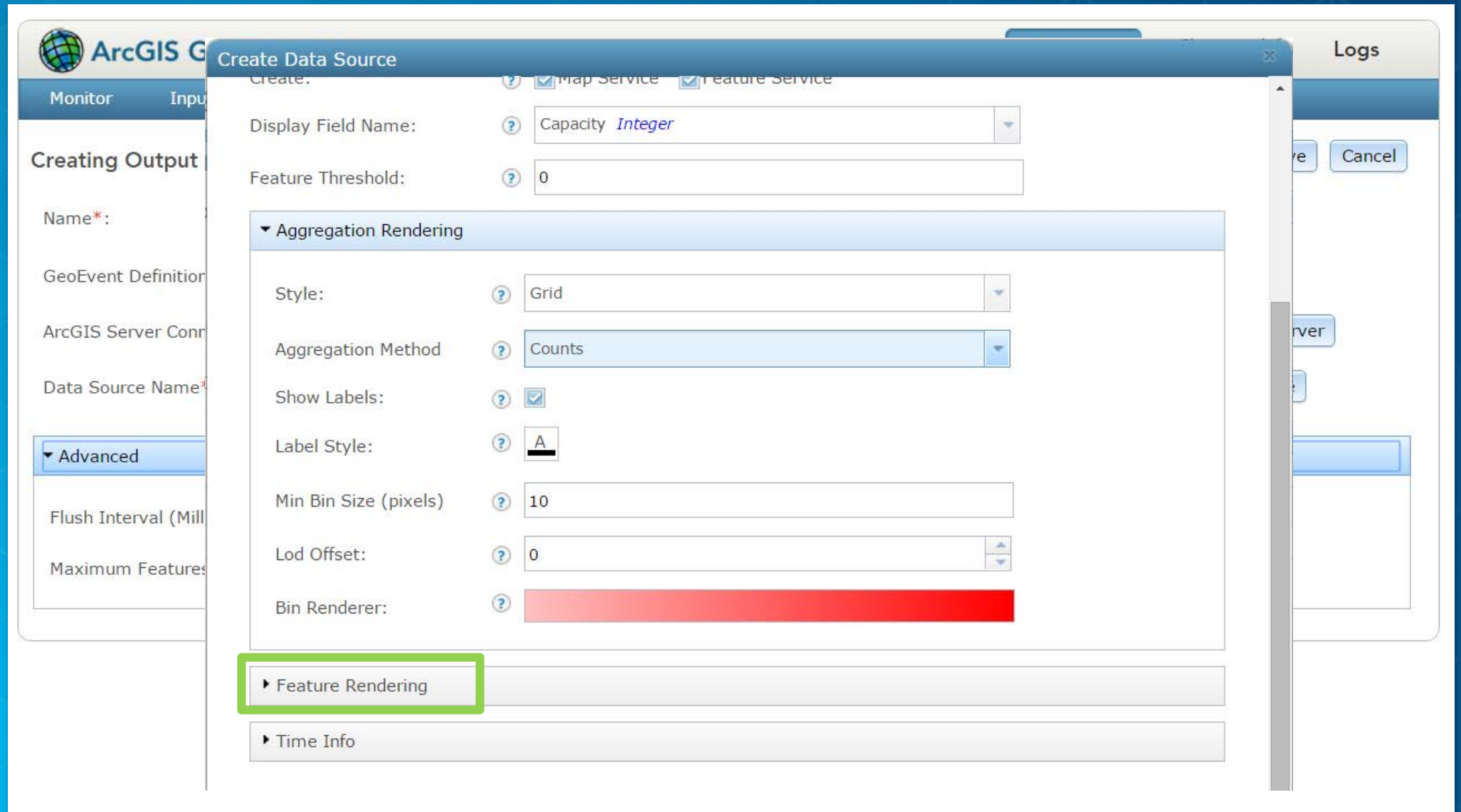
10.4

Add a Feature to an ArcGIS Spatiotemporal Big Data Store



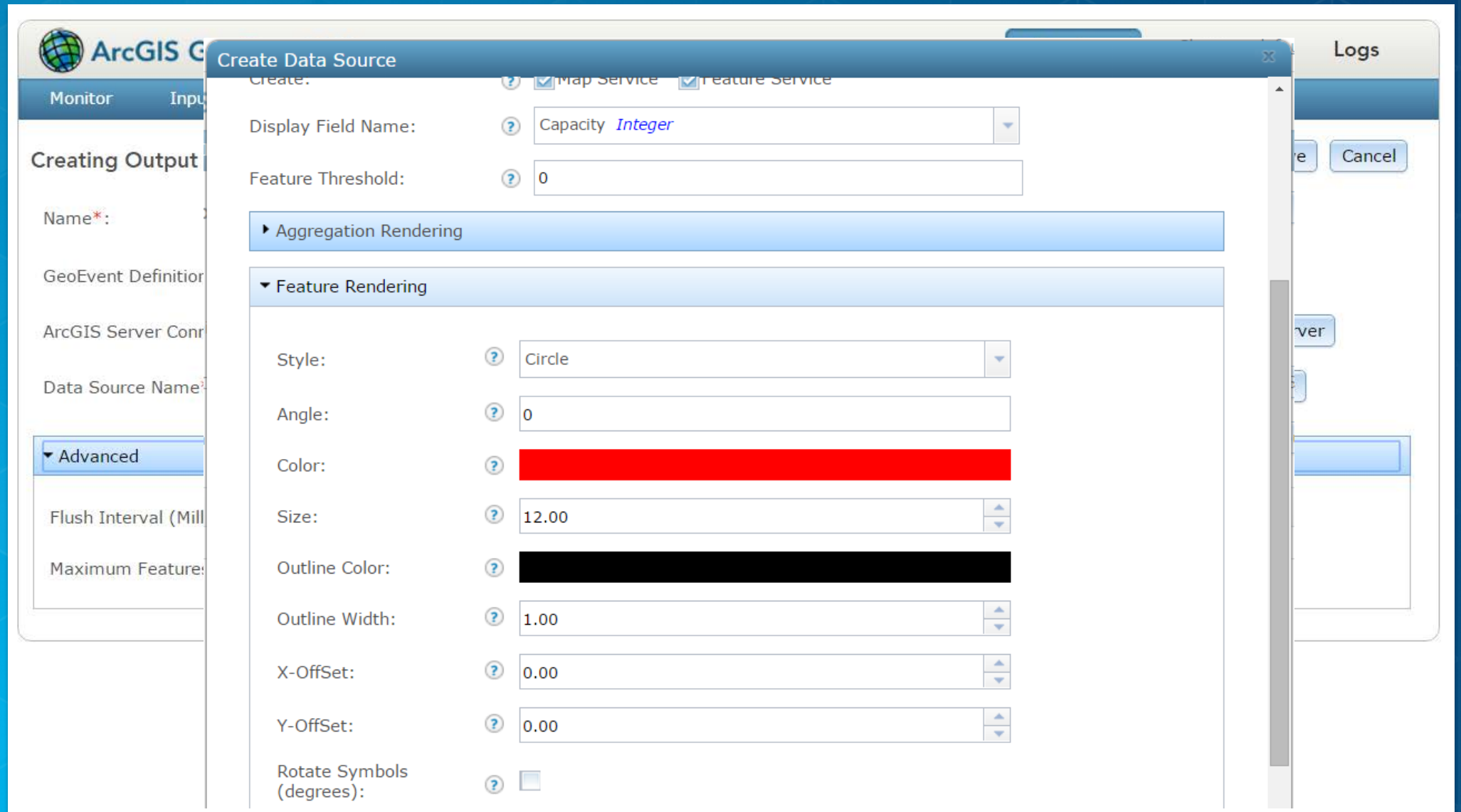
Storing Observation Data

Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

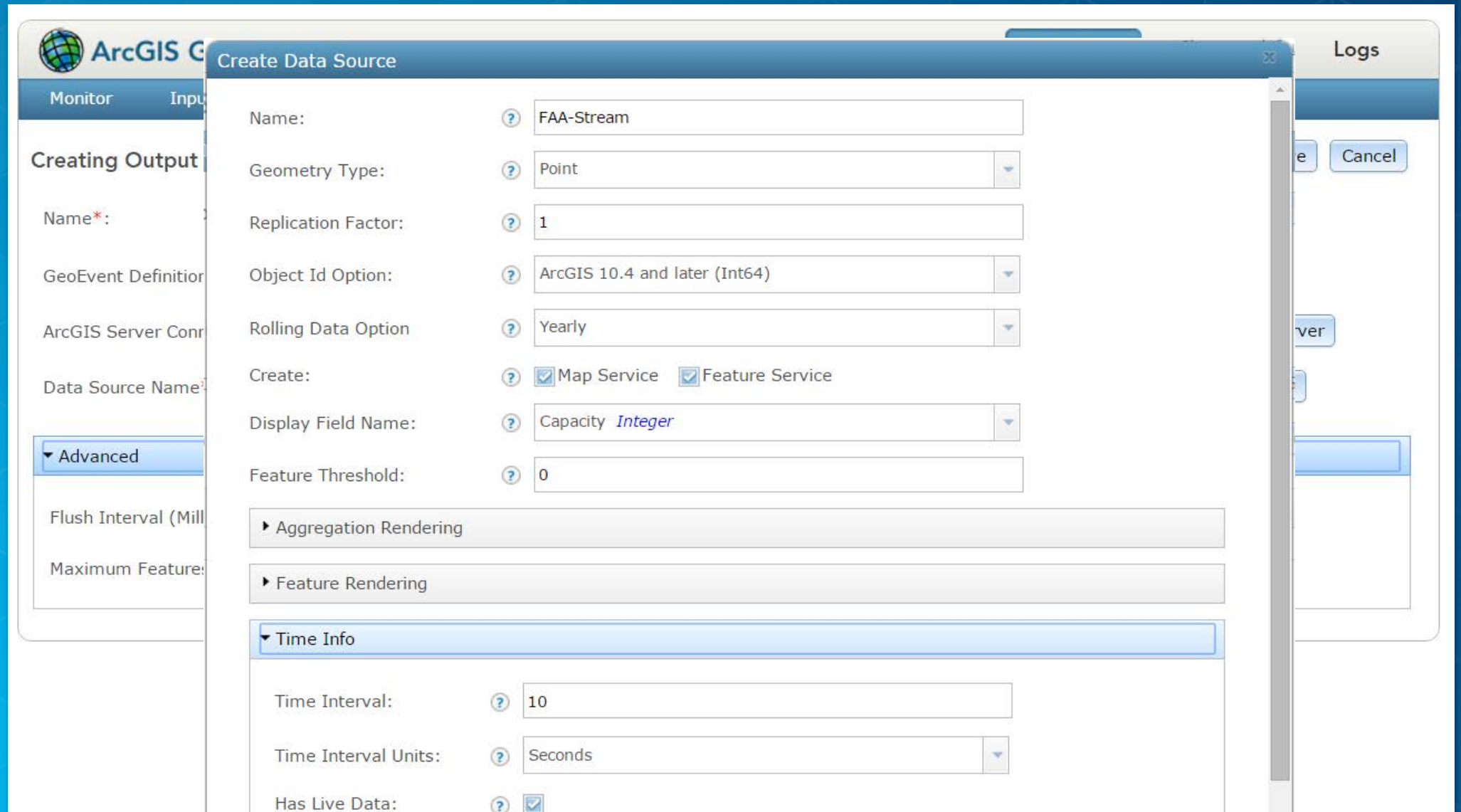
Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

10.4

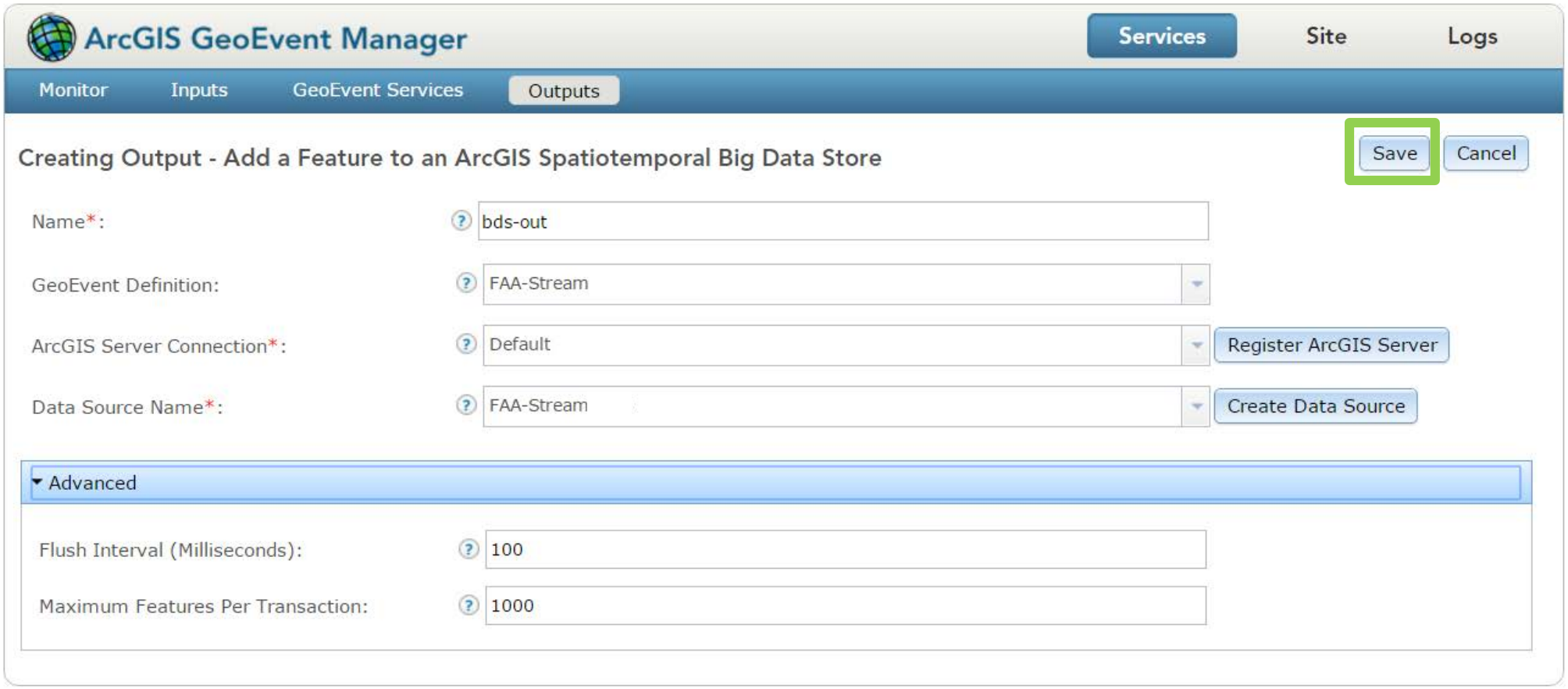
Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

10.4

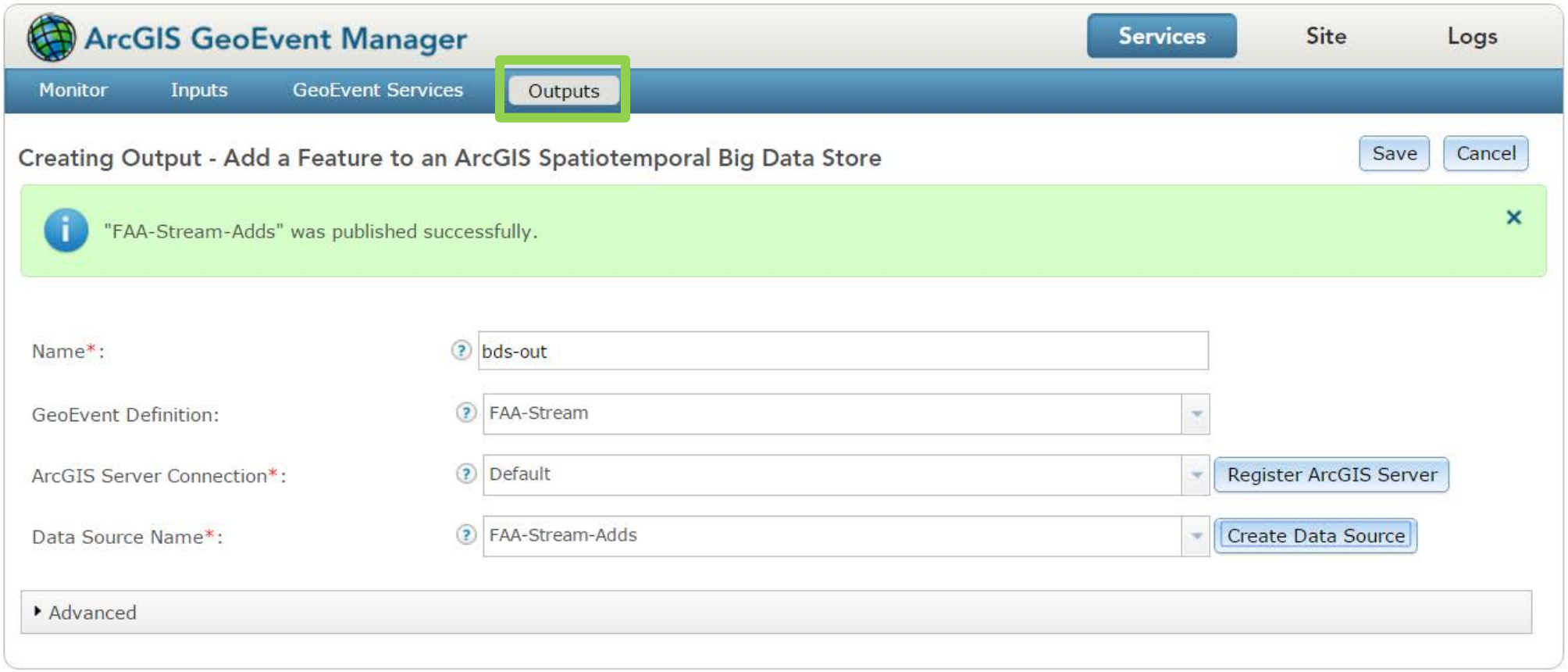
Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

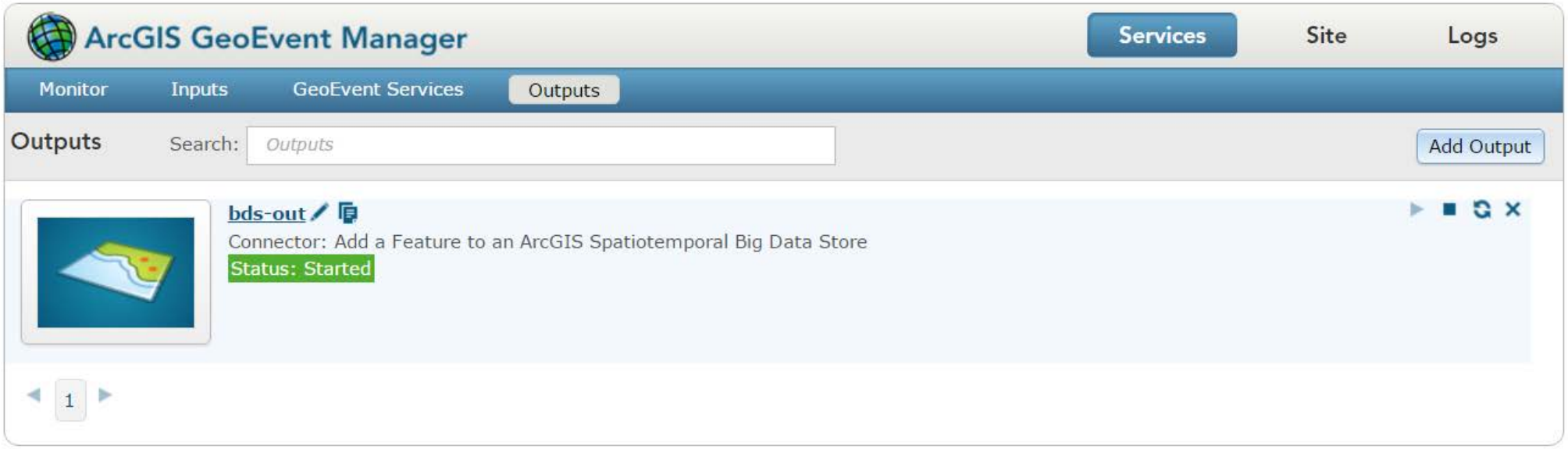
10.4

Add a Feature to an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

Add a Feature to an ArcGIS Spatiotemporal Big Data Store

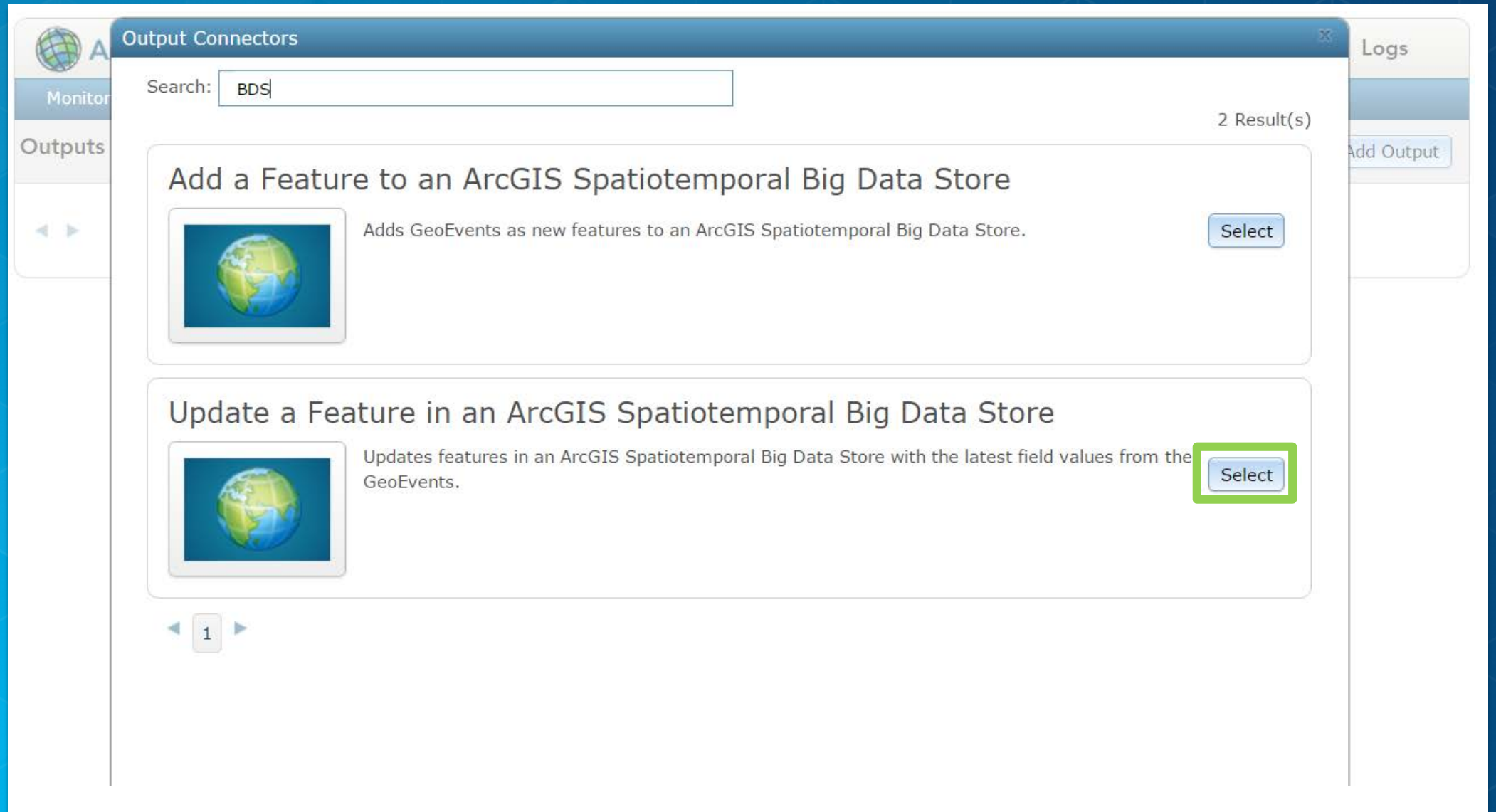


5b. Storing Observation Data

Update a Feature

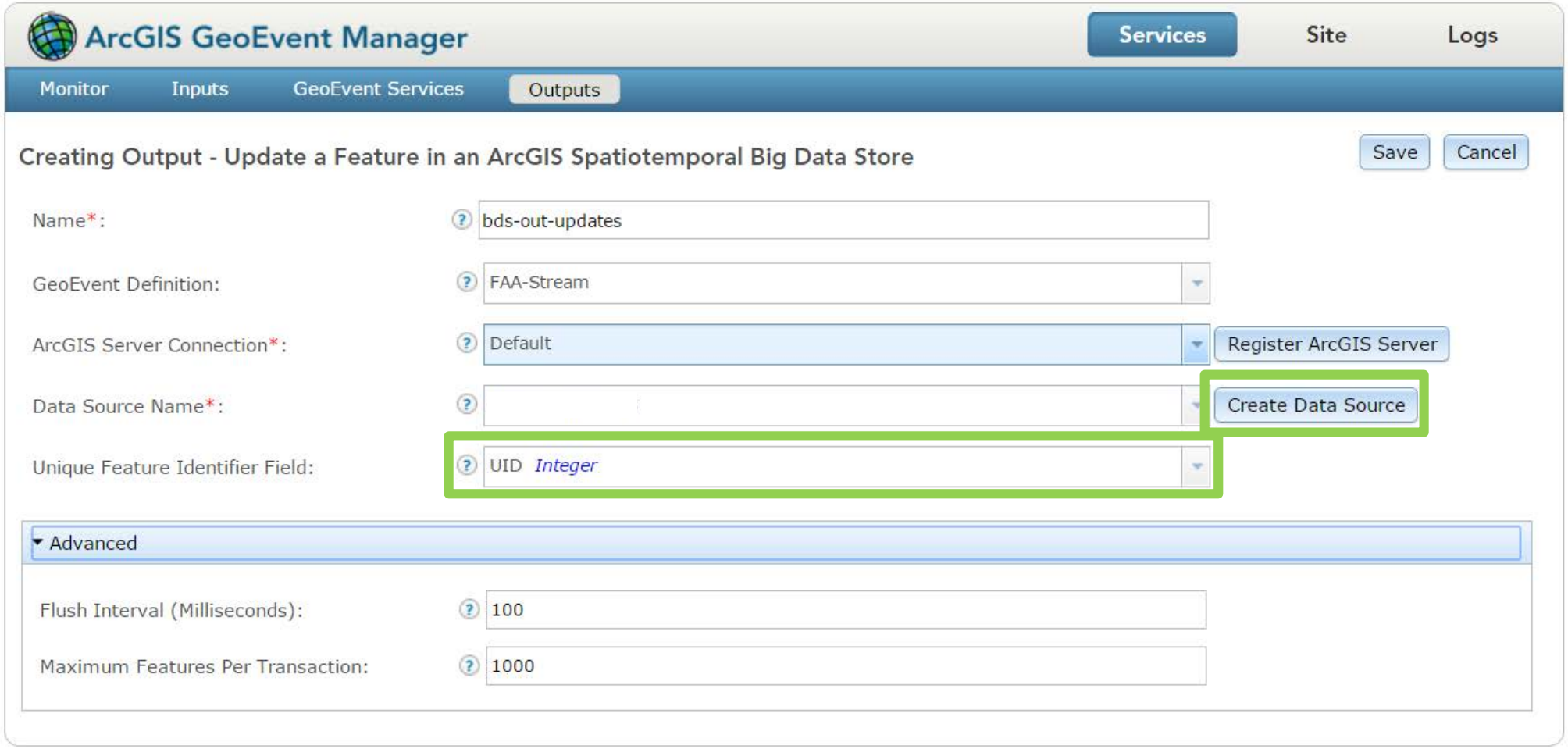
Storing Observation Data

Update a feature in an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

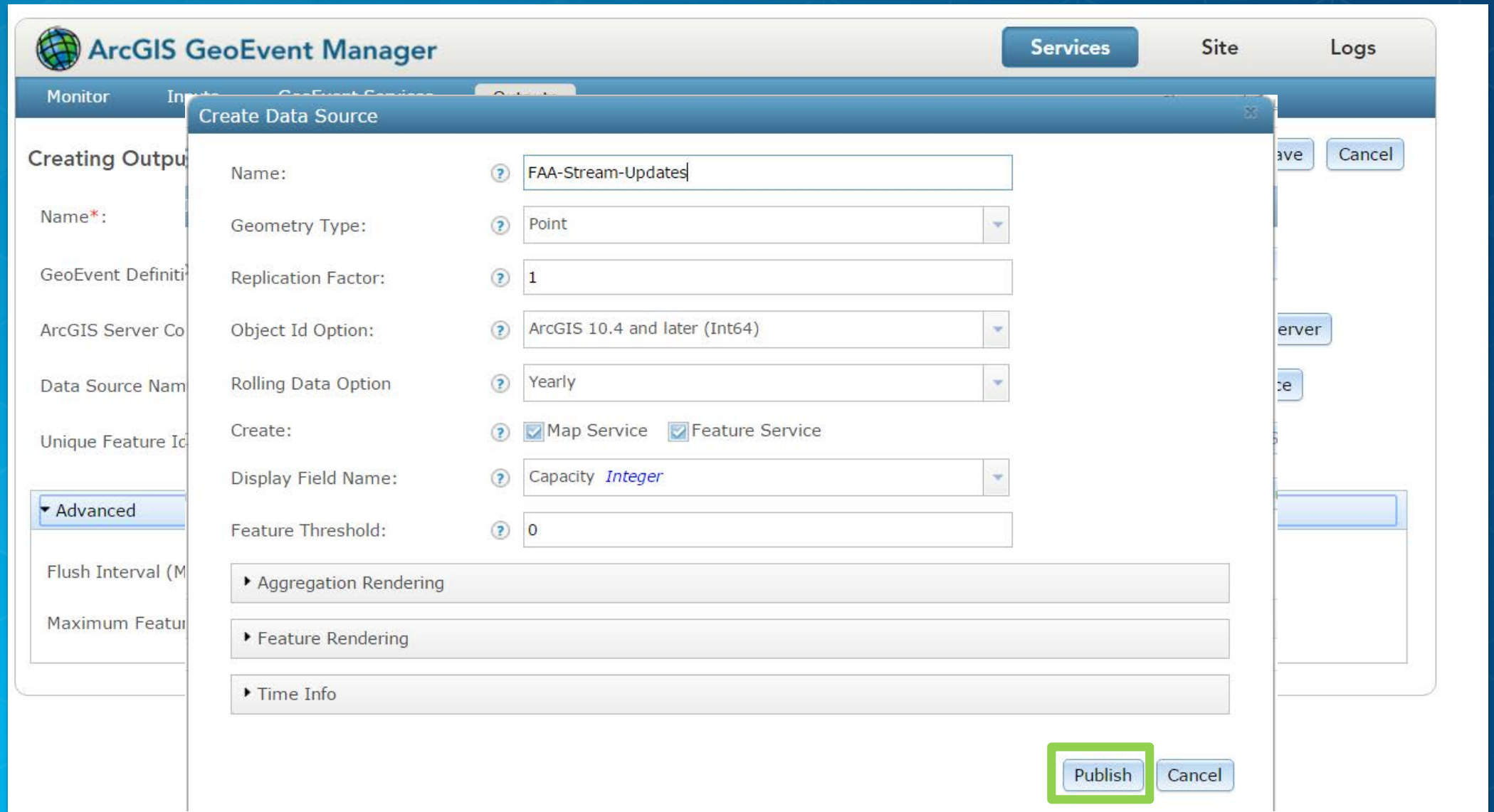
Update a feature in an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

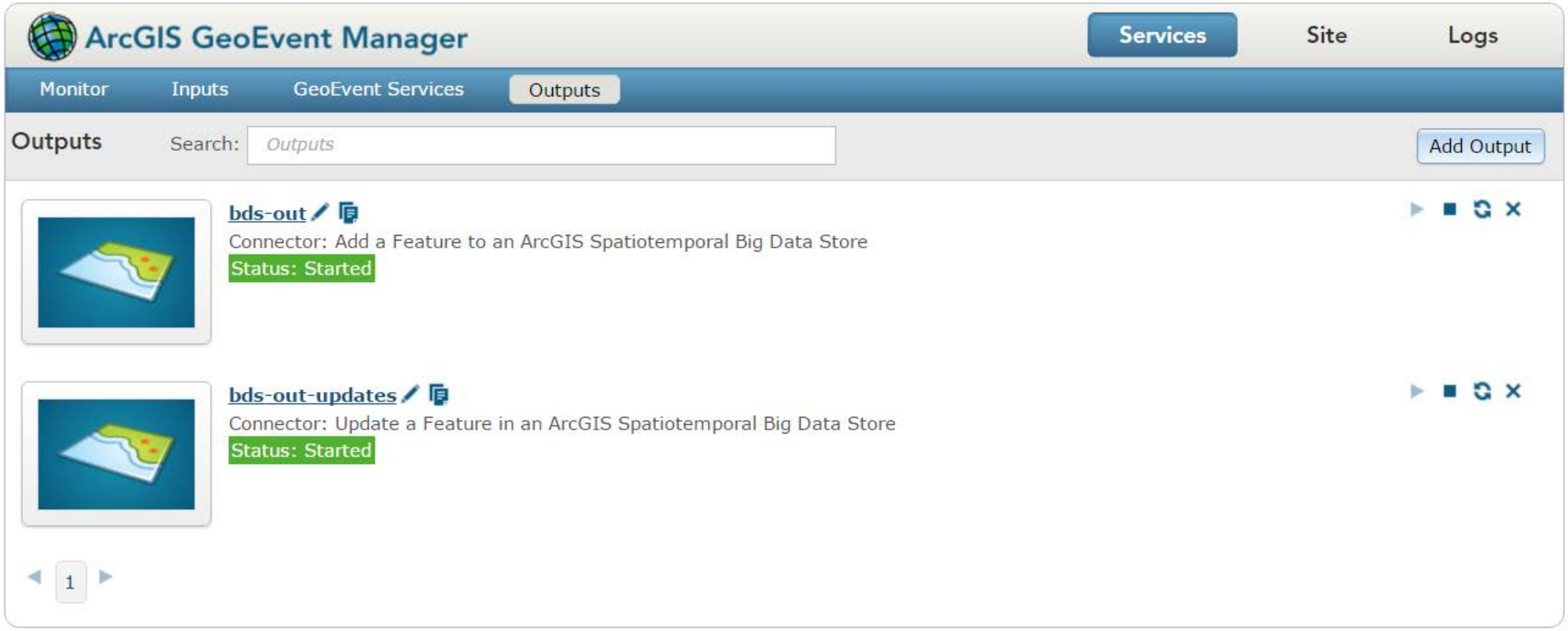
10.4

Update a feature in an ArcGIS Spatiotemporal Big Data Store



Storing Observation Data

Update a feature in an ArcGIS Spatiotemporal Big Data Store



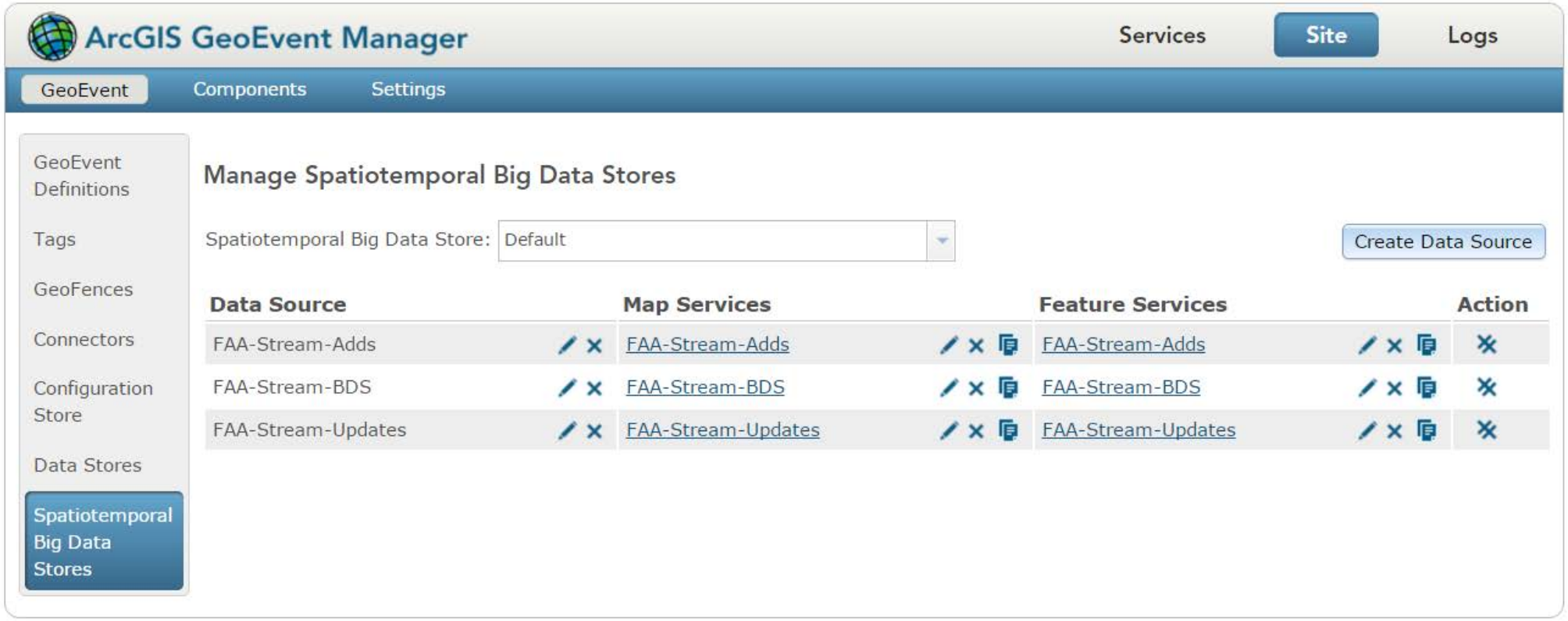
5c. Storing Observation Data

Managing a Big Data Store

Storing Observation Data

Managing a Big Data Store

10.4

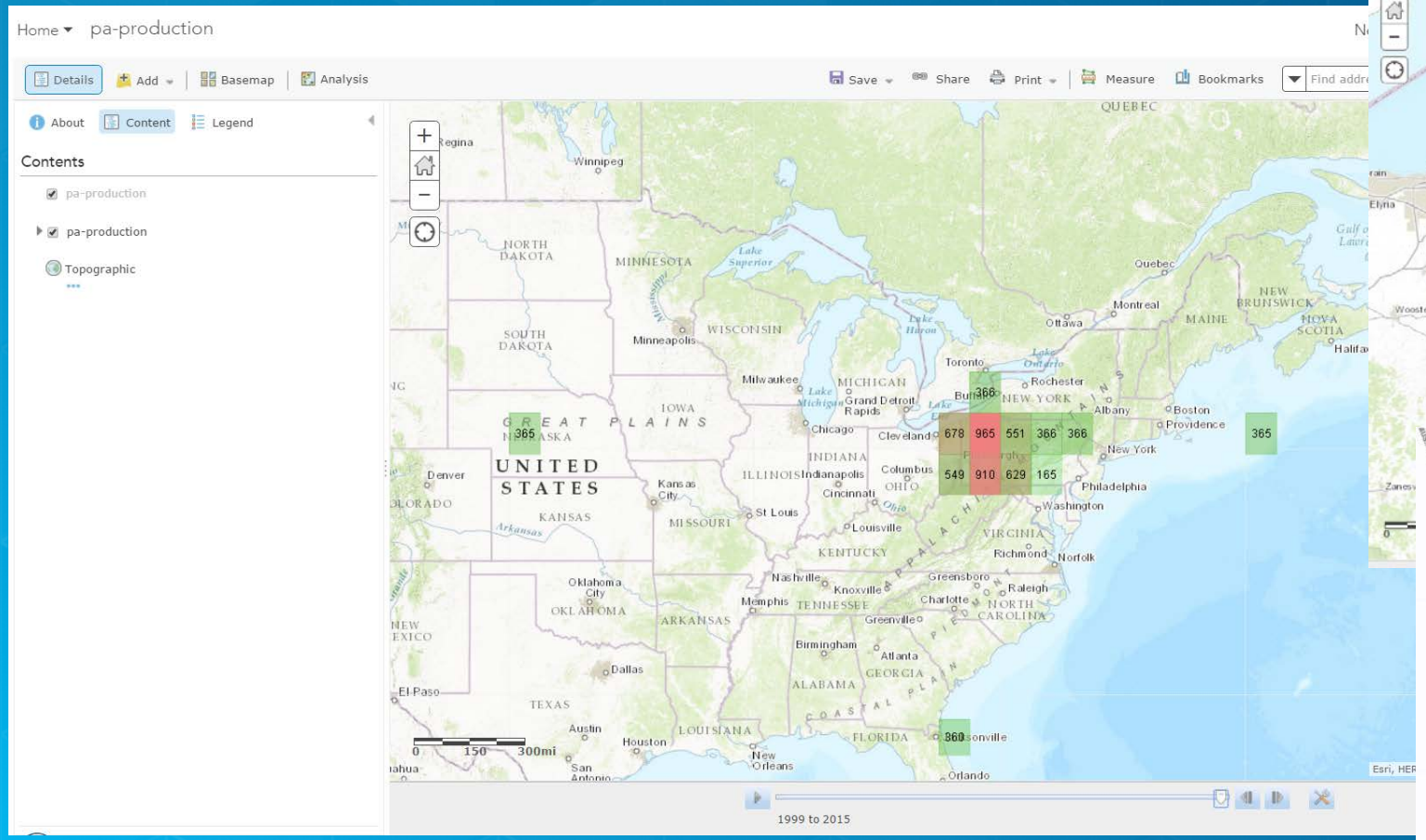


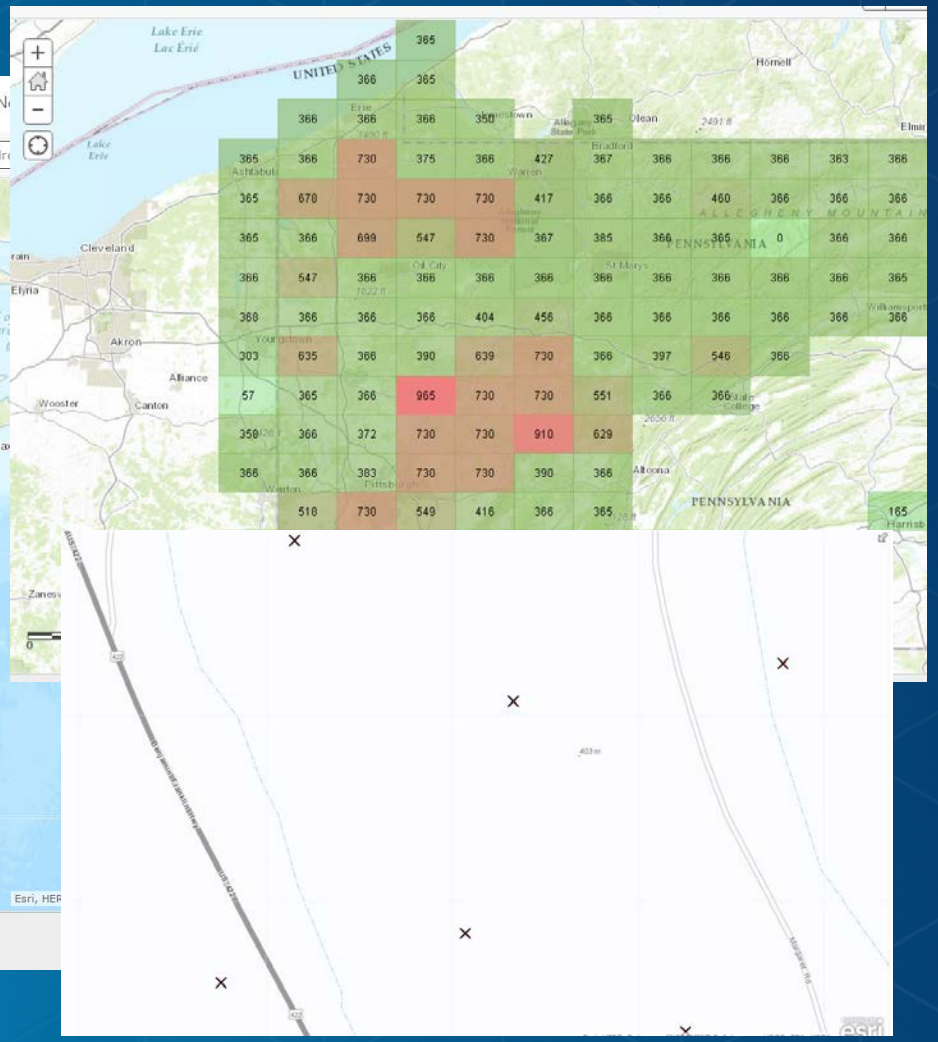
Demo

10.4

Visualizing Observation Data

Map & Feature Service reading from Big Data Store





6. A Deeper Dive

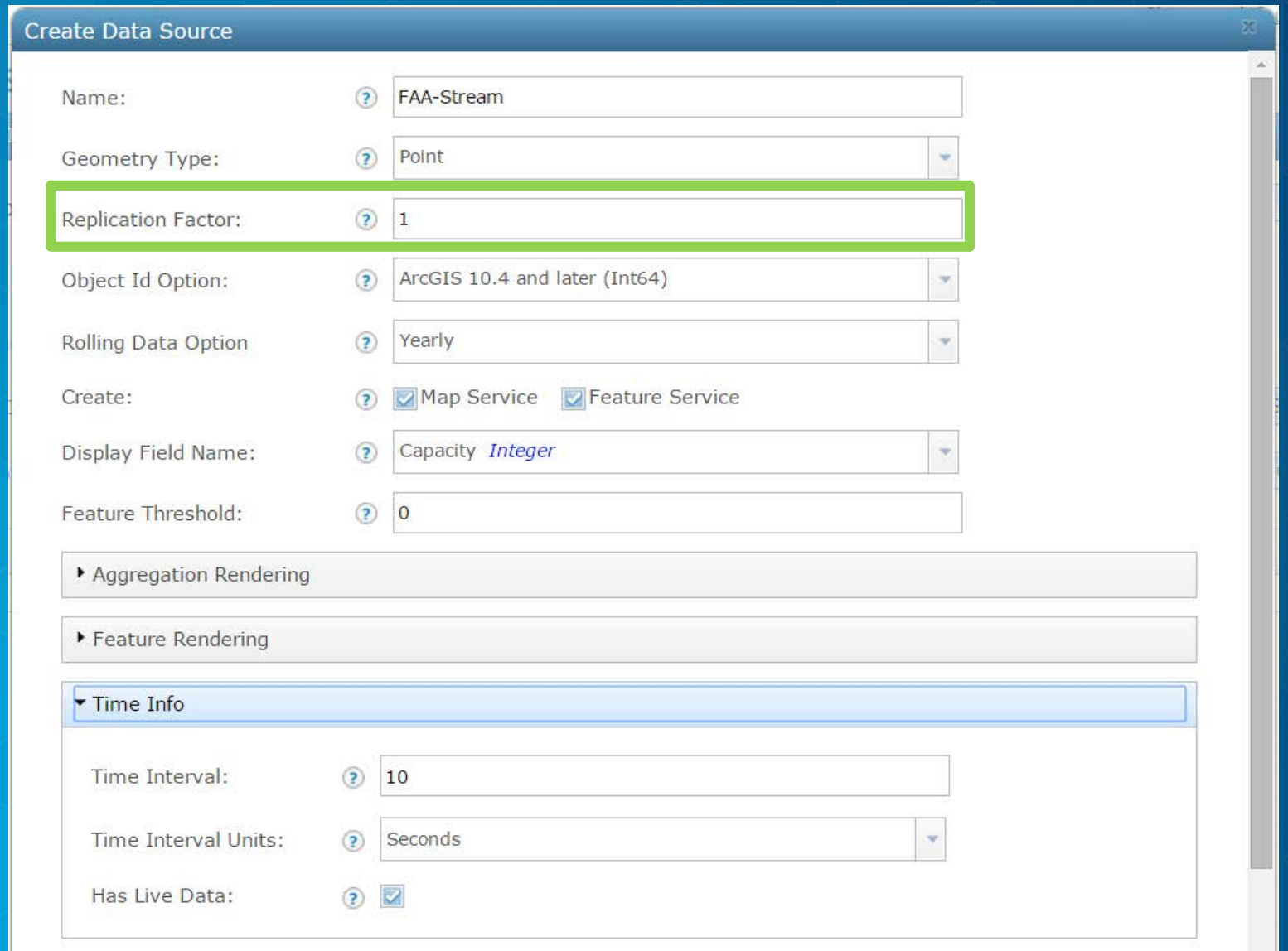
6a. A Deeper Dive

Data Reliability

Data Reliability

Replication Factor

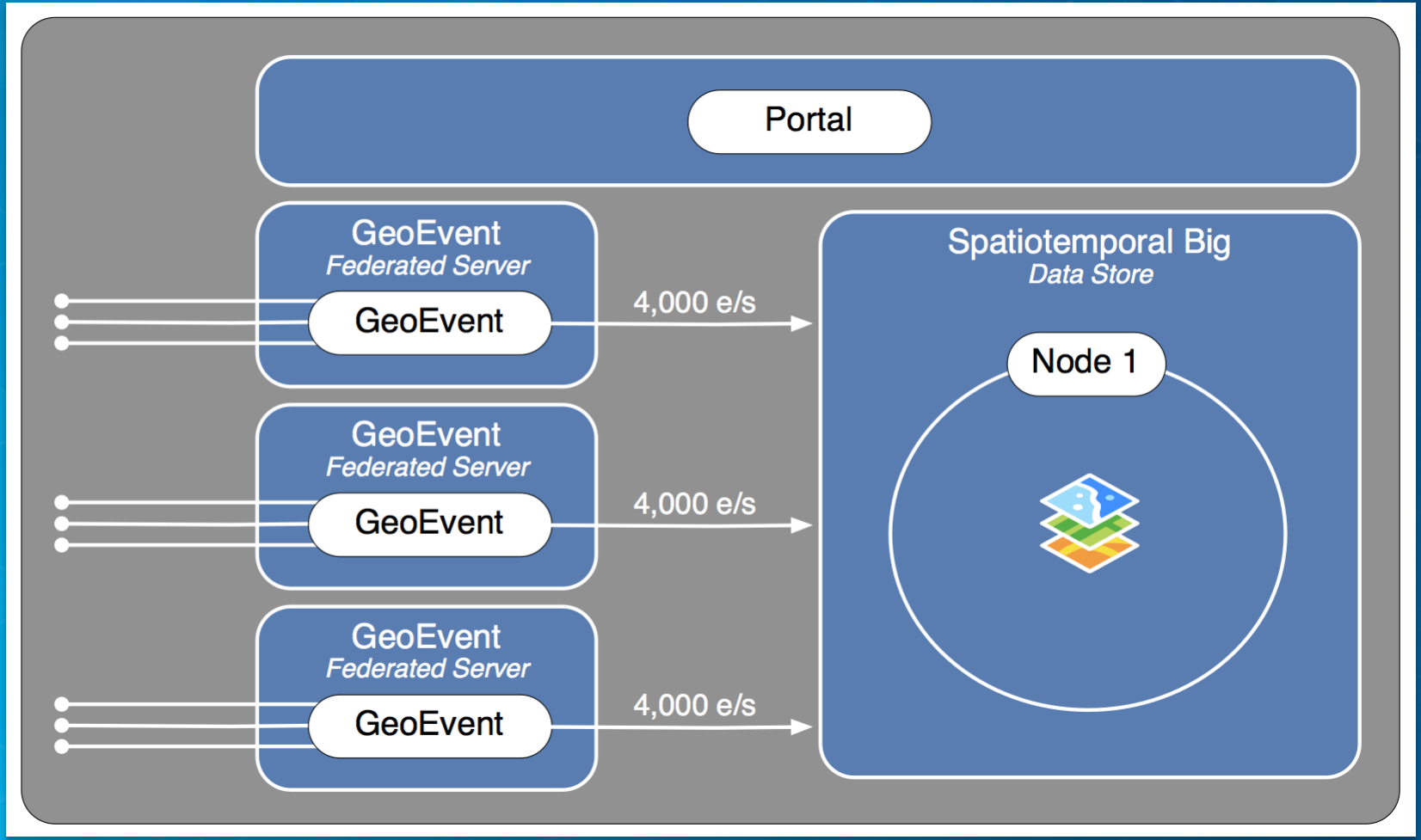
10.4



Data Reliability

Replication Factor

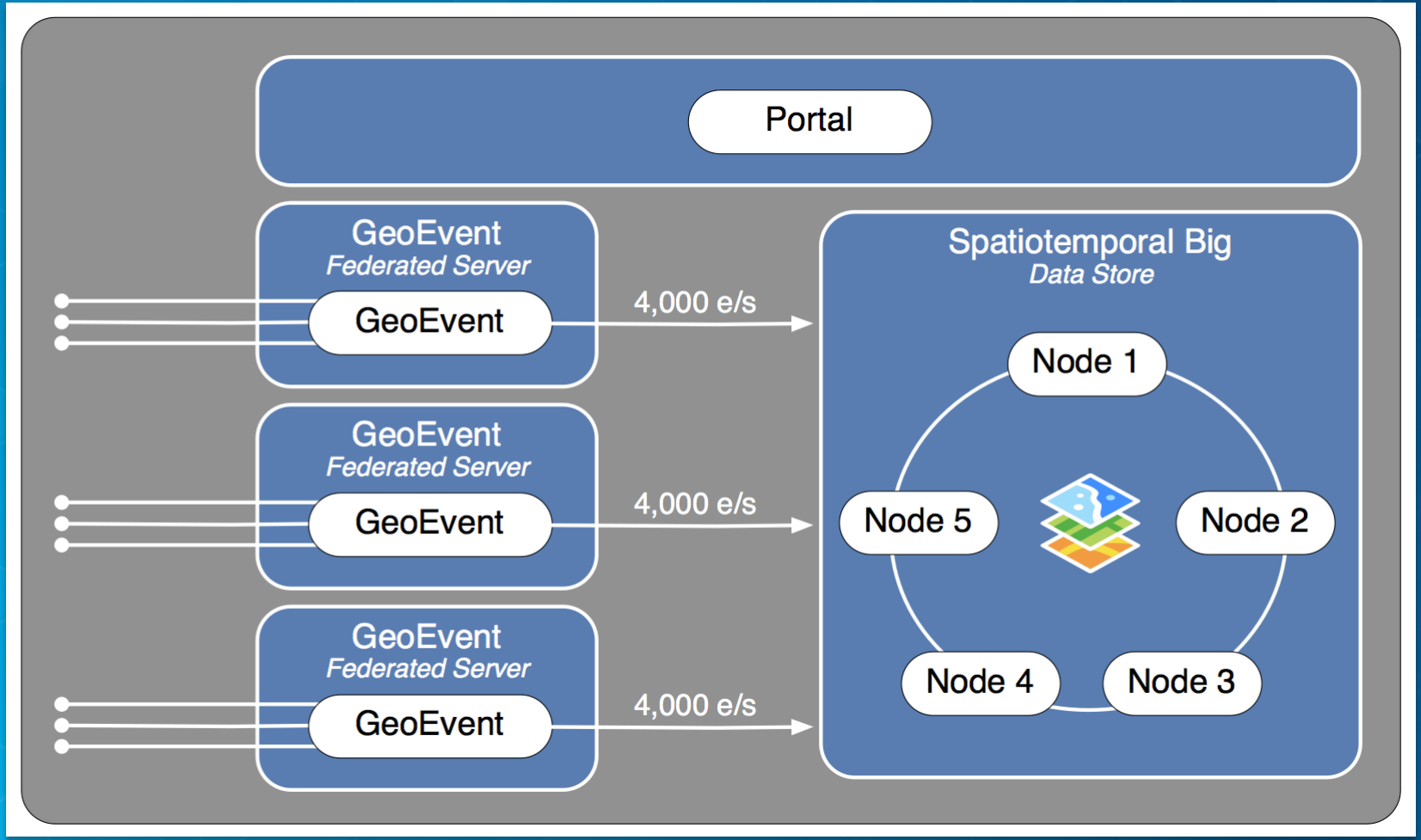
10.4



Data Reliability

Replication Factor

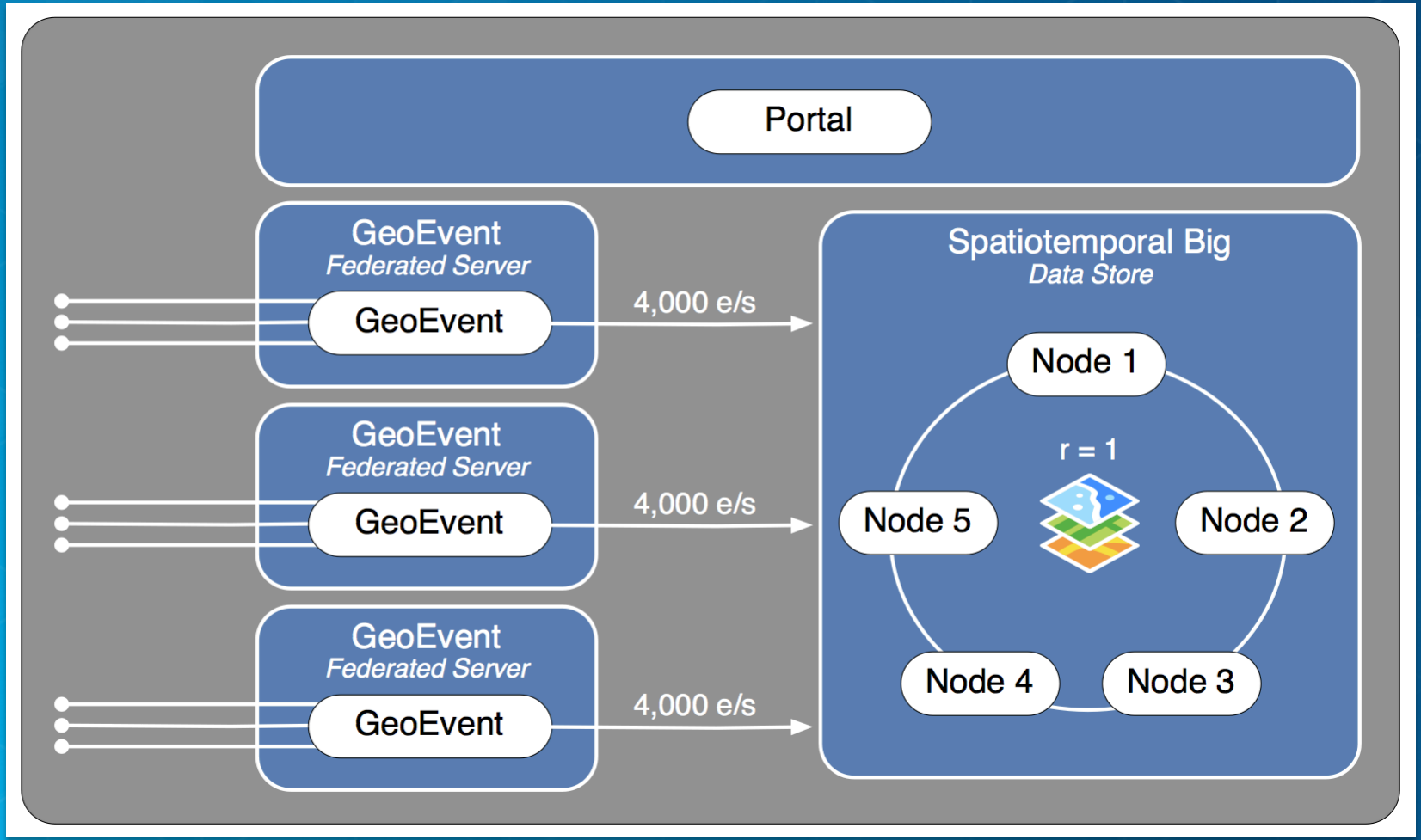
10.4



Data Reliability

Replication Factor

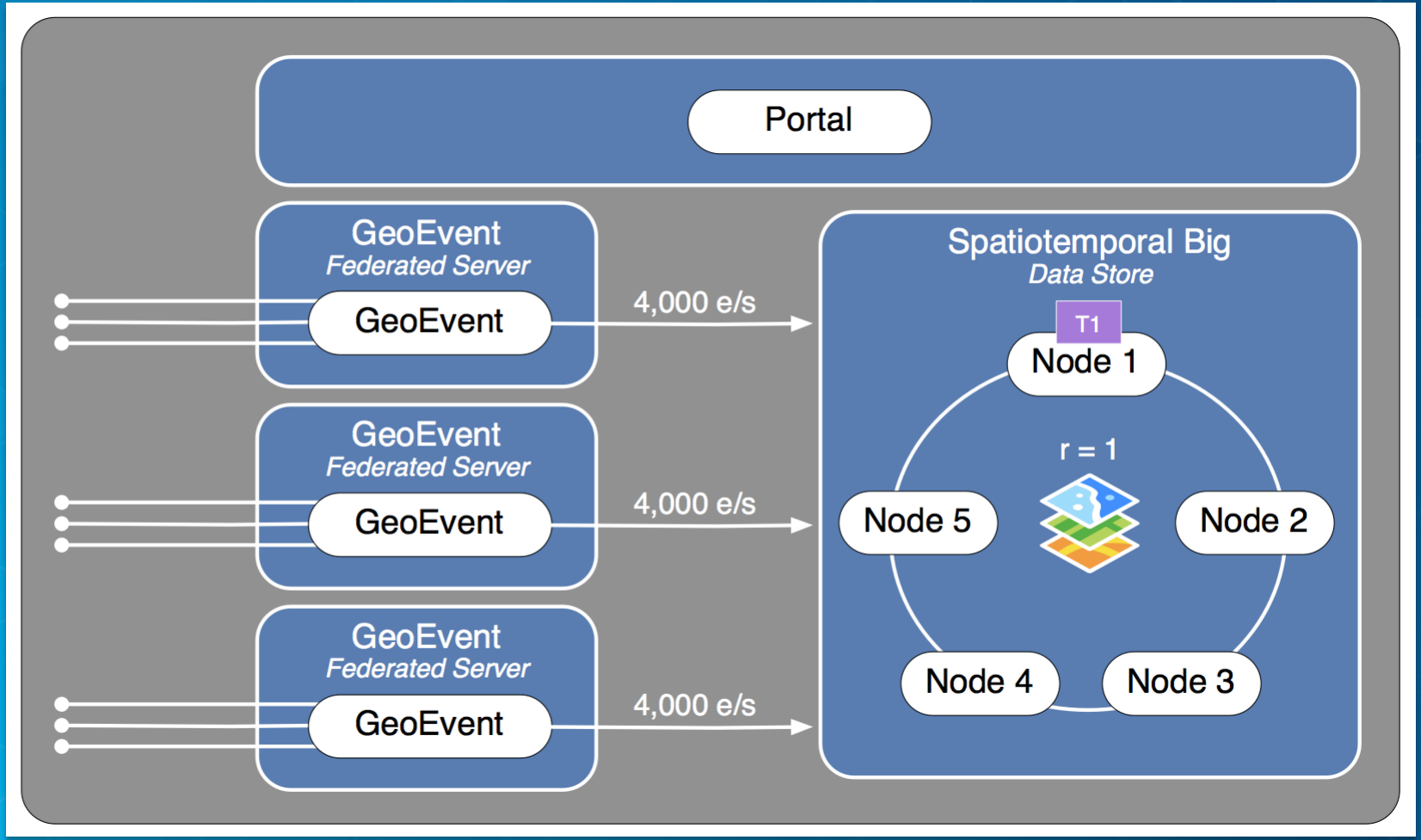
10.4



Data Reliability

Replication Factor

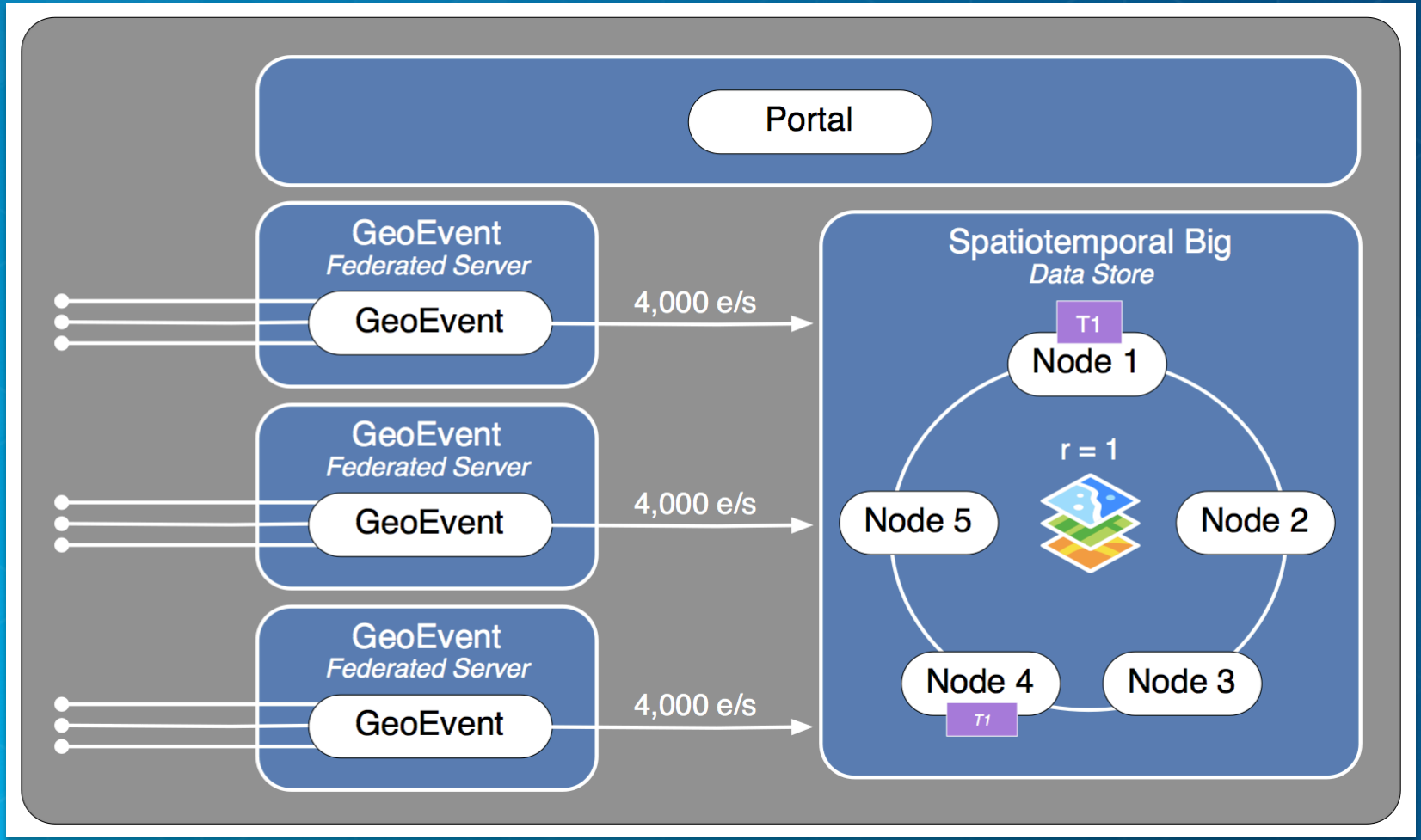
10.4



Data Reliability

Replication Factor

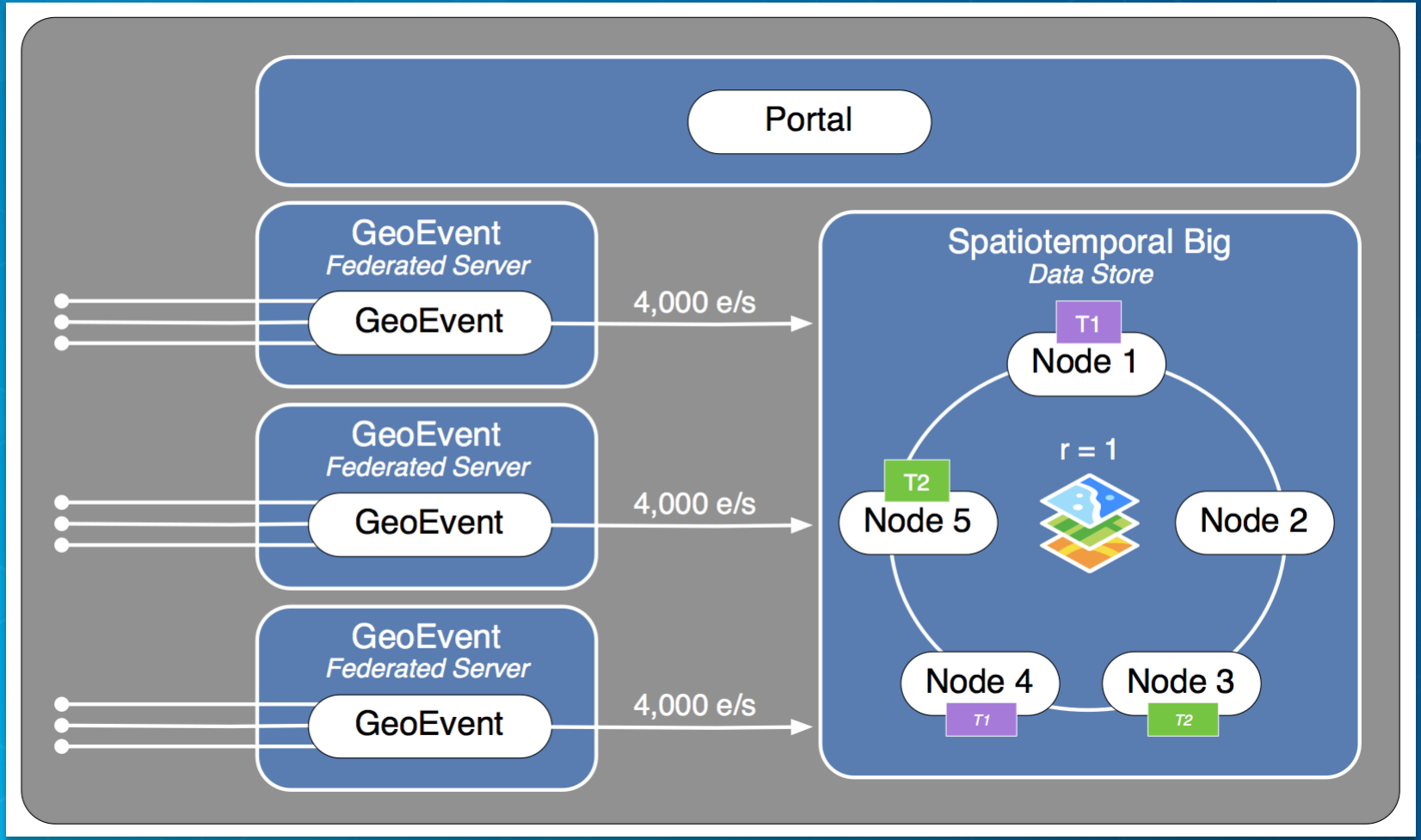
10.4



Data Reliability

Replication Factor

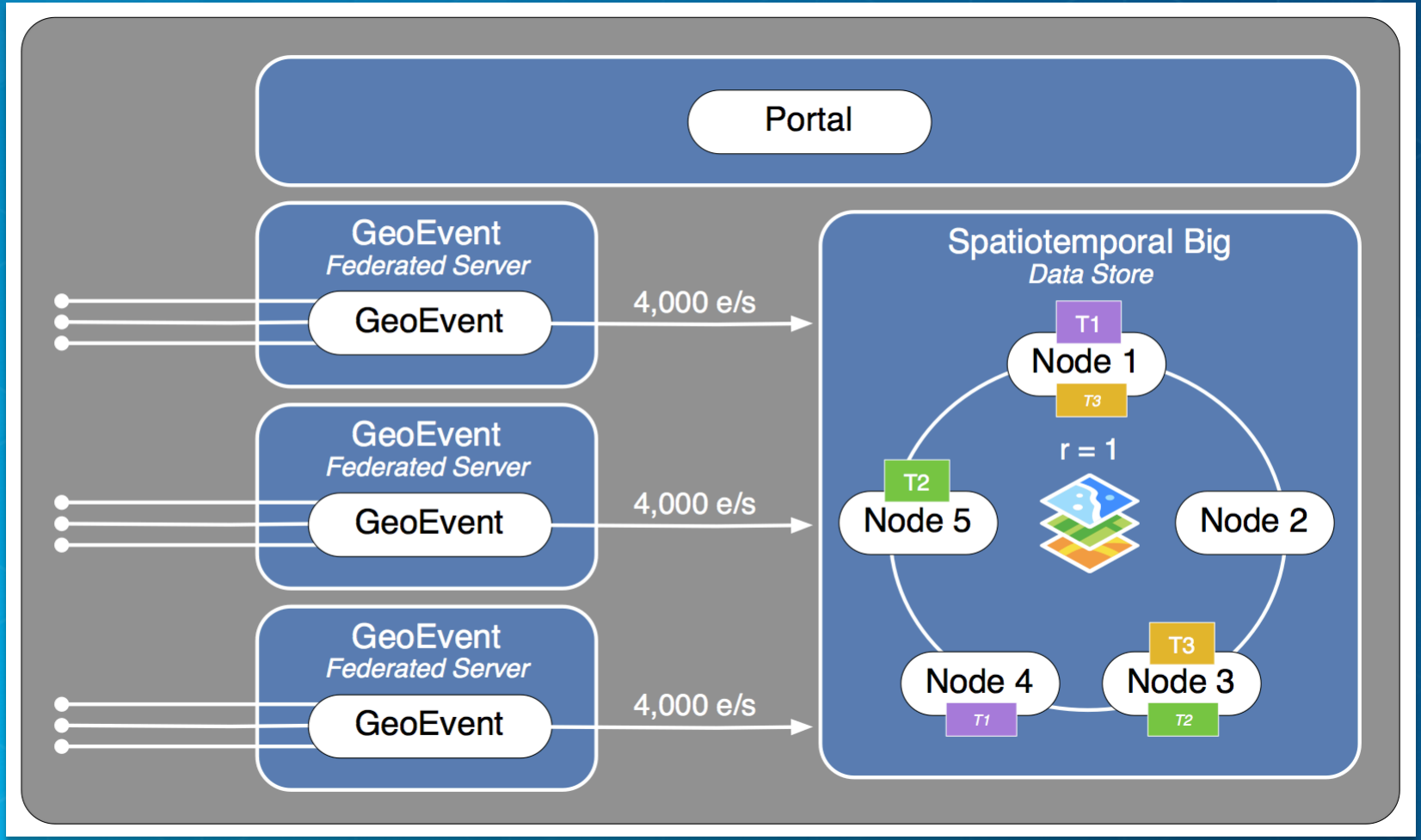
10.4



Data Reliability

Replication Factor

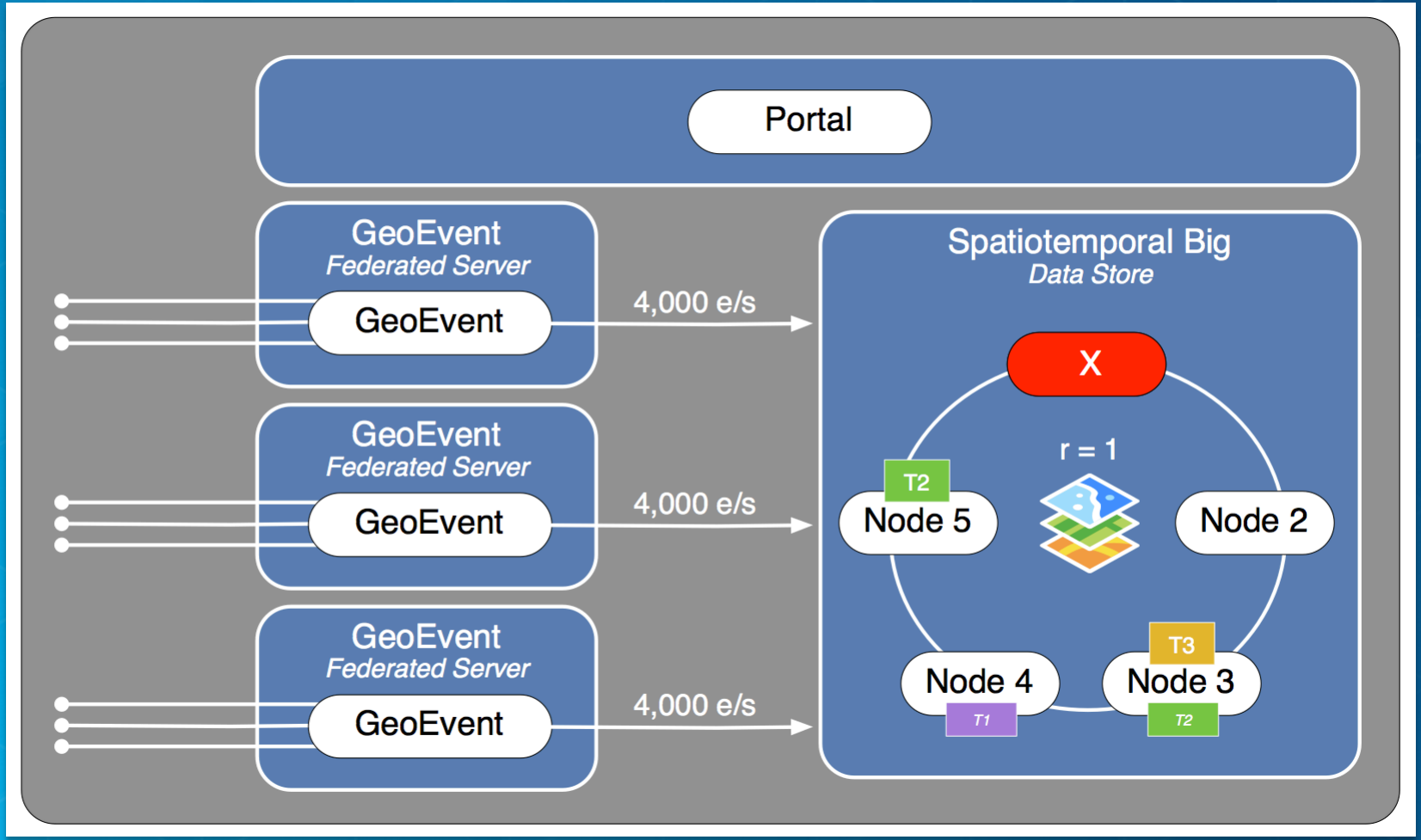
10.4



Data Reliability

Node Failure

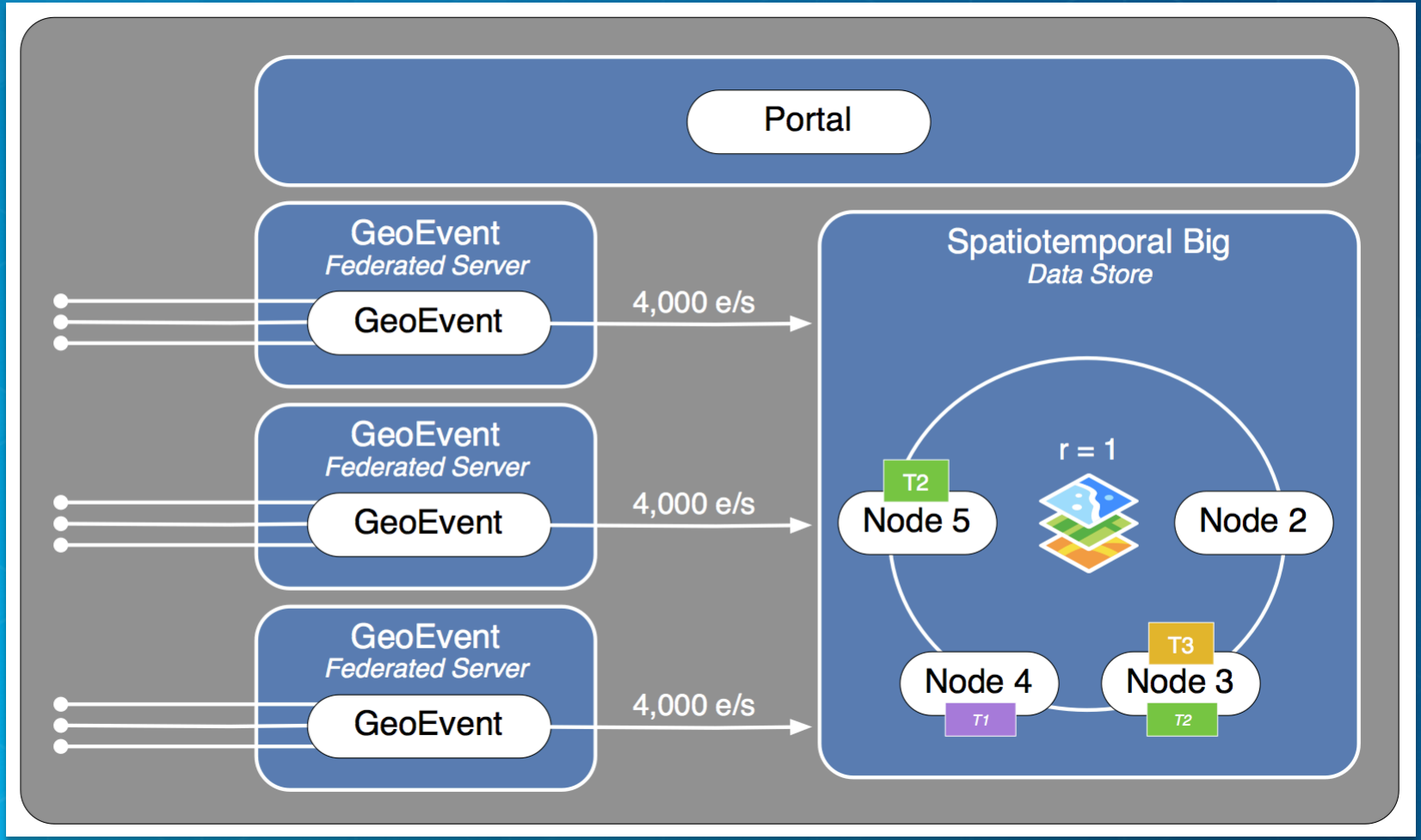
10.4



Data Reliability

Auto-rebalancing of data upon node membership changes, + or -, in the Big Data Store

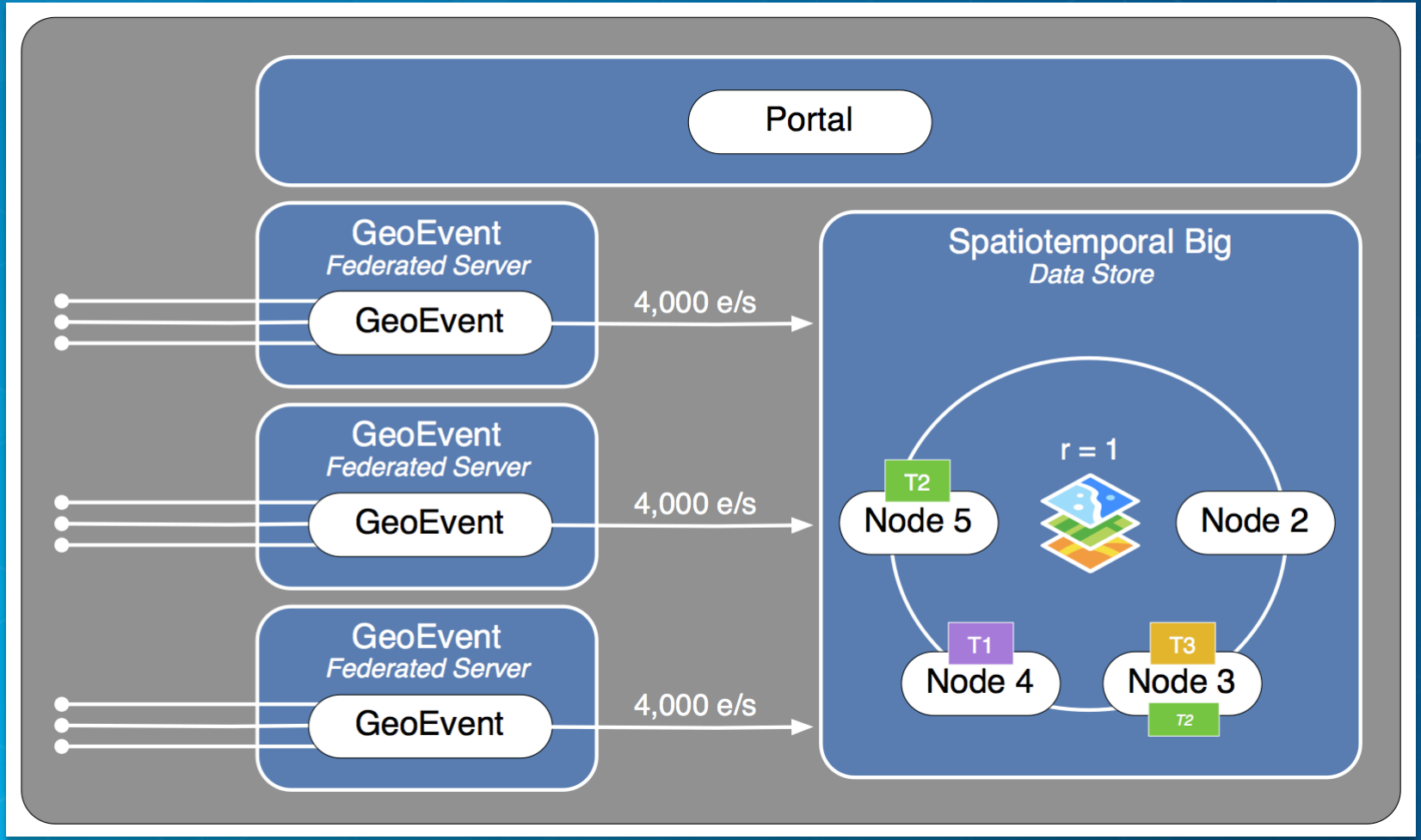
10.4



Data Reliability

Auto-rebalancing of data upon node membership changes, + or -, in the Big Data Store

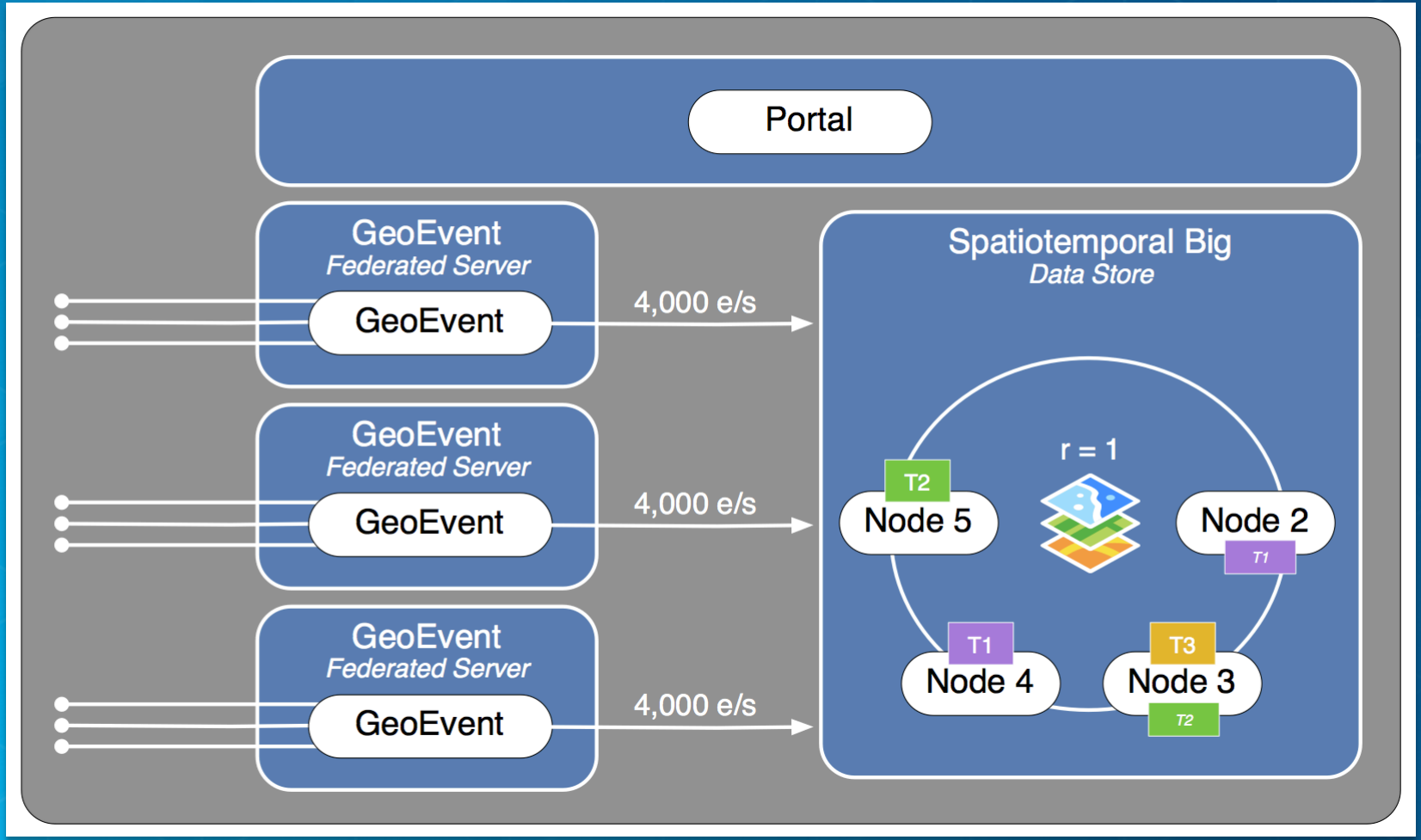
10.4



Data Reliability

Auto-rebalancing of data upon node membership changes, + or -, in the Big Data Store

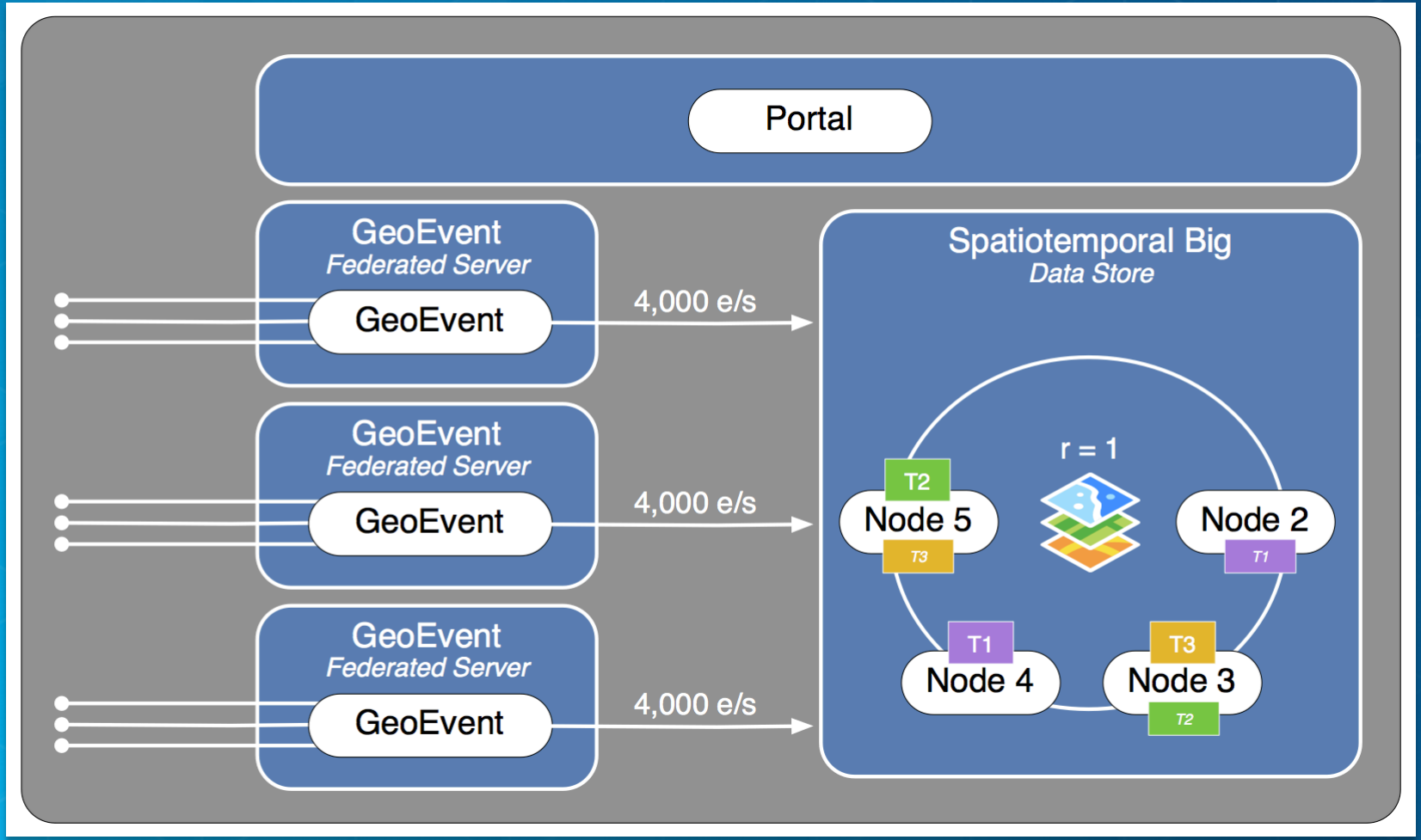
10.4



Data Reliability

Auto-rebalancing of data upon node membership changes, + or -, in the Big Data Store

10.4



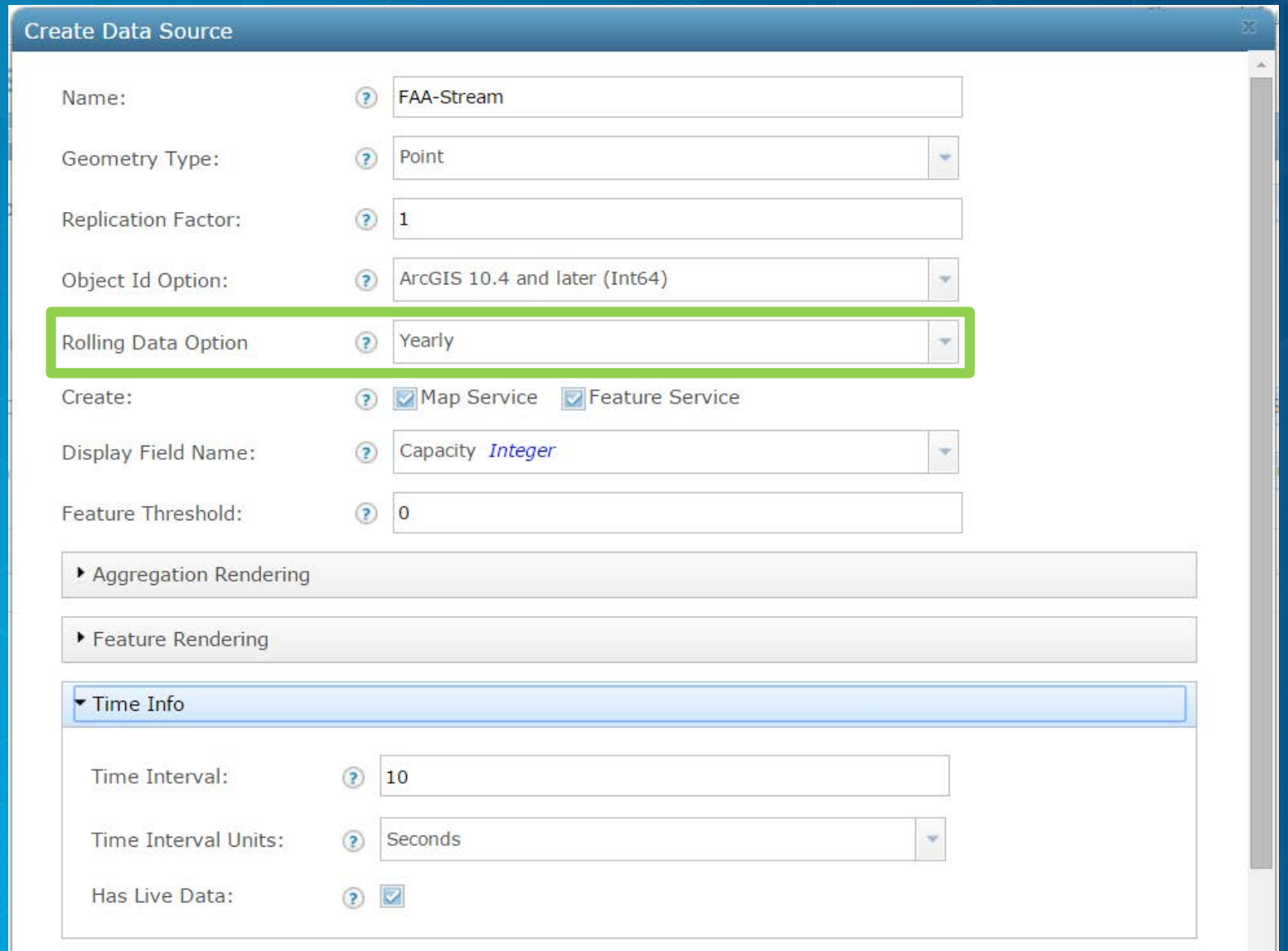
6b. A Deeper Dive

Rolling Data Option

Rolling Data

Enabling fast access & queries

10.4



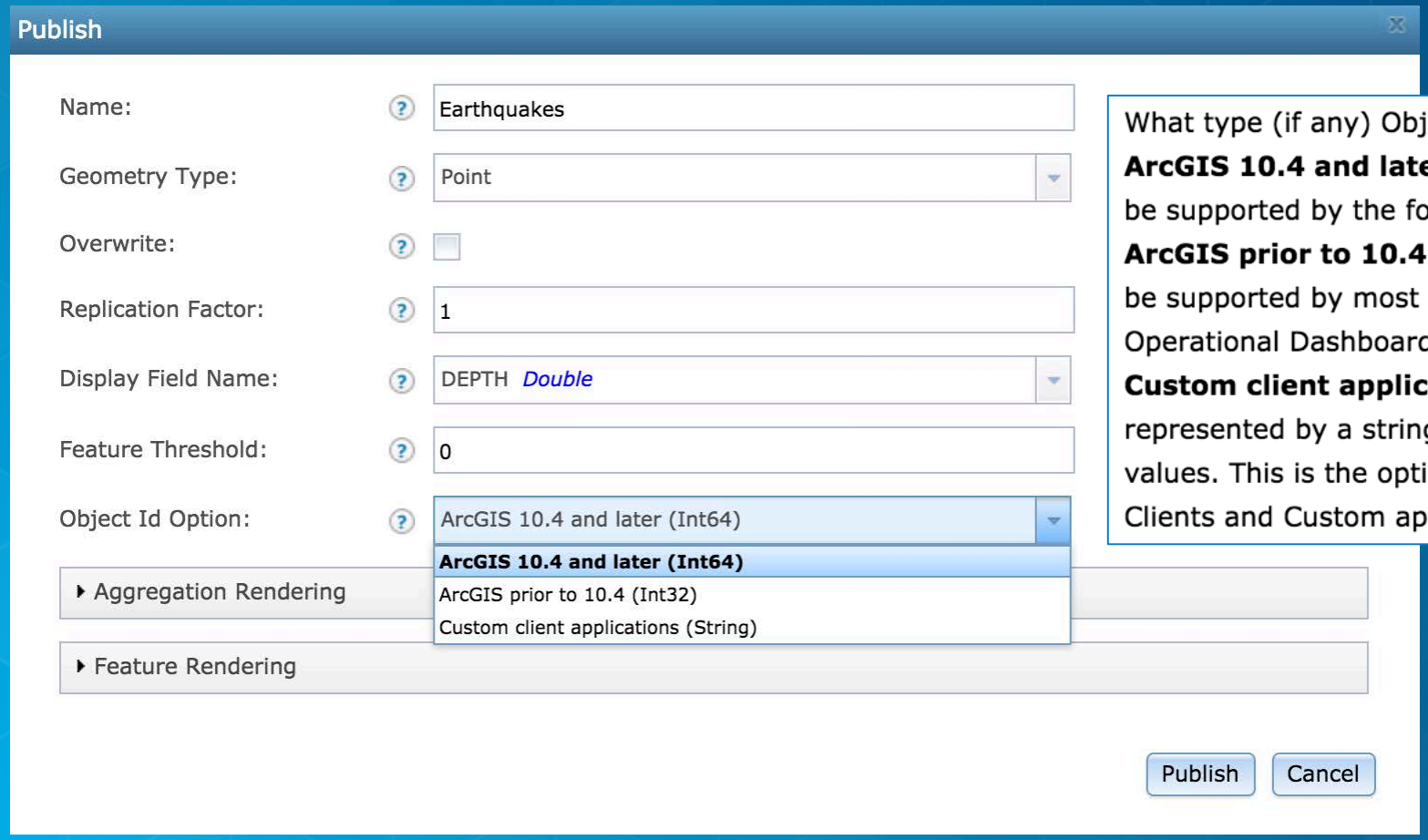
6c. A Deeper Dive

Observations and ObjectID

Observations and ObjectID

Choosing an ObjectID Option: Int32, Int64, or UniqueStringID

10.4



What type (if any) Object Id to generated.

ArcGIS 10.4 and later (Int64): Generates objectIds which will be supported by the following clients:

ArcGIS prior to 10.4 (Int32): Generates objectIds which will be supported by most clients including: ArcMap, ArcGIS Pro, Operational Dashboard, etc.

Custom client applications (String): The Object Id is represented by a string field which holds unique identifier string values. This is the optional write throughput setting. Javascript Clients and Custom applications will support this approach.

Observations and ObjectID

Choosing an ObjectID Option: Int32, Int64, or UniqueStringID

10.4

	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

	events per day	Int32	Int64 (signed)
1,000 e/s	86,400,000	25 days	292,472,000 years
10,000 e/s	864,000,000	2.5 days	29,247,200 years
100,000 e/s	8,640,000,000	6 hours	2,924,720 years
1,000,000 e/s	86,400,000,000	36 minutes	292,472 years
10,000,000 e/s	864,000,000,000	4 minutes	29,248 years

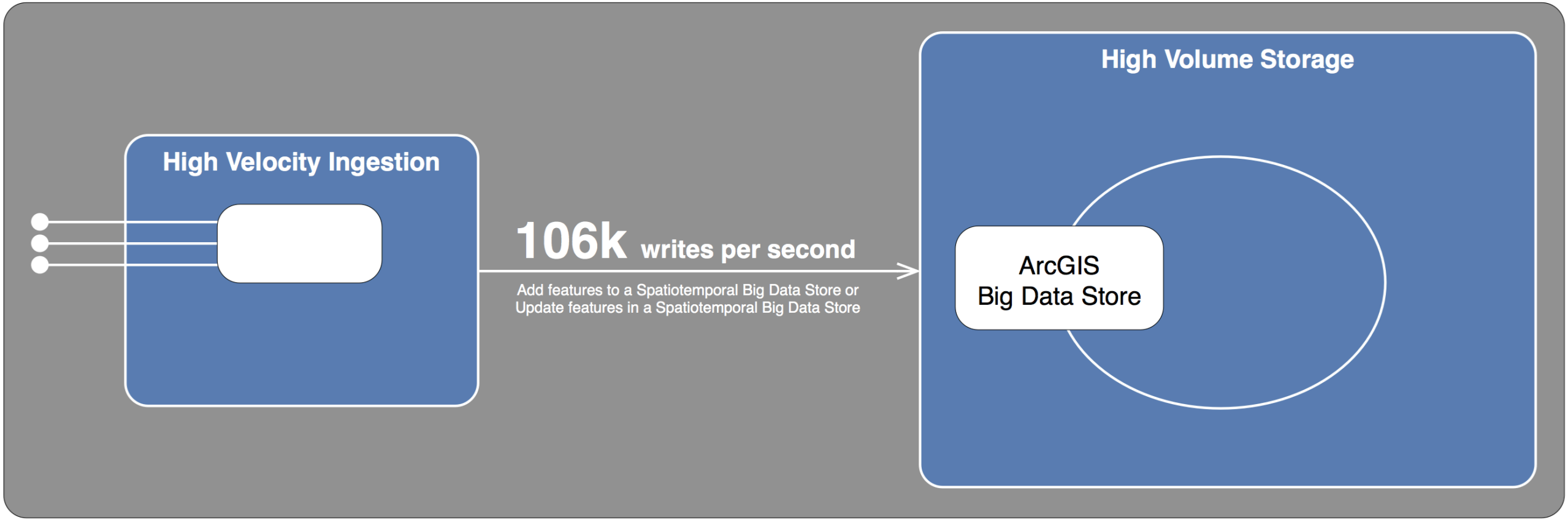
7. Capacity Planning

7a. Capacity Planning

Big Data Store Performance

Big Data Store Performance

ArcGIS 10.4 Benchmarks *c4.2xlarge (Windows 2012 Server R2): 8 vCPU, 15 GiB, 100GB local SSD, 1,000 Mbps EBS*



Storage Write Throughput	1 node
ArcGIS Big Data Store	106k
ArcGIS Relational Data Store	0.2k

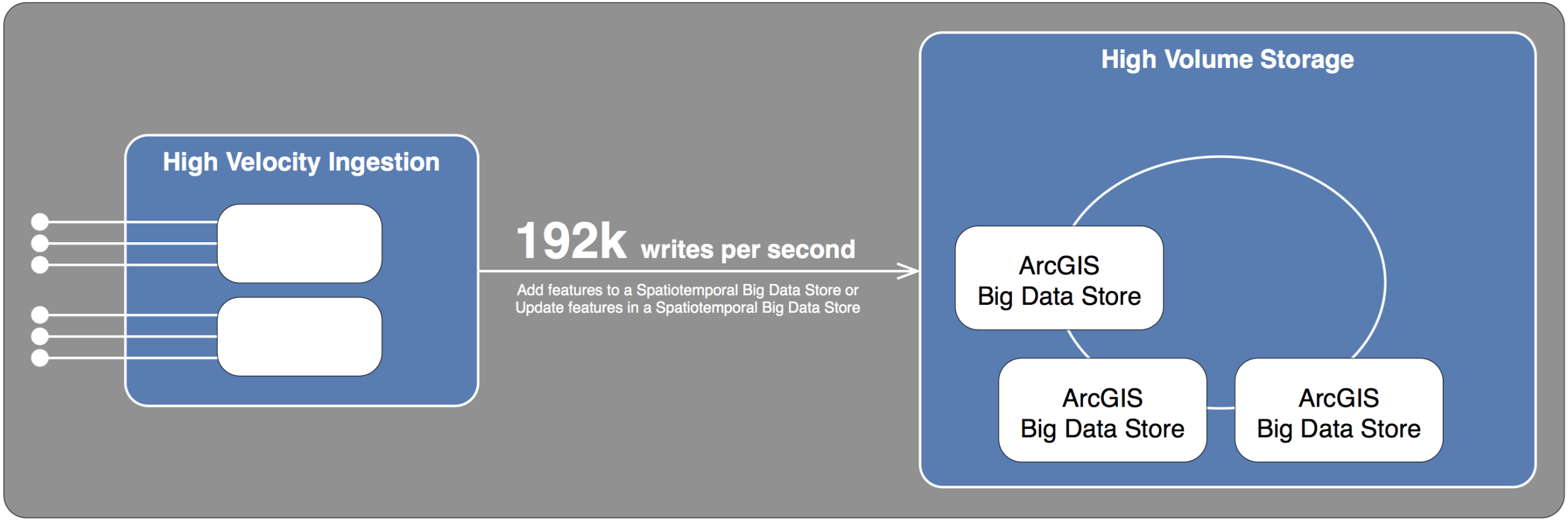
Big Data Store Performance

ArcGIS 10.4 Benchmarks *c4.2xlarge (Windows 2012 Server R2): 8 vCPU, 15 GiB, 100GB local SSD, 1,000 Mbps EBS*

Storage Write Throughput	1 node	2 node
ArcGIS Big Data Store	106k	143k
ArcGIS Relational Data Store	0.2k	n/a

Big Data Store Performance

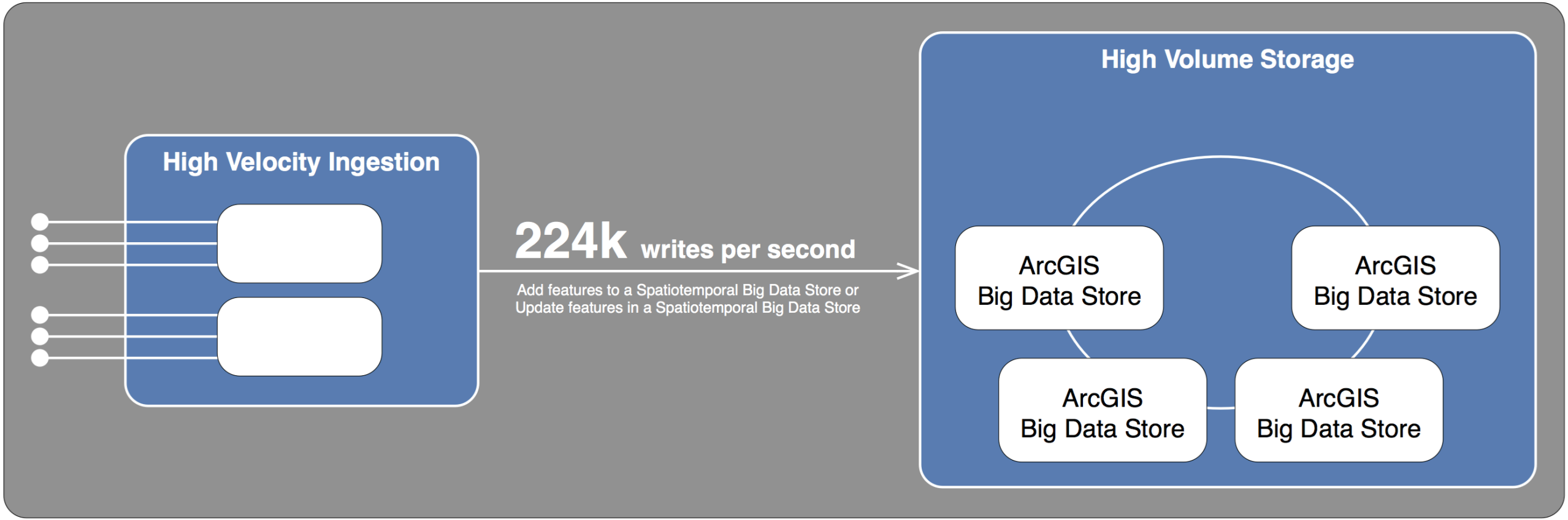
ArcGIS 10.4 Benchmarks *c4.2xlarge (Windows 2012 Server R2): 8 vCPU, 15 GiB, 100GB local SSD, 1,000 Mbps EBS*



Storage Write Throughput	1 node	2 node	3 node
ArcGIS Big Data Store	106k	143k	192k
ArcGIS Relational Data Store	0.2k	n/a	n/a

Big Data Store Performance

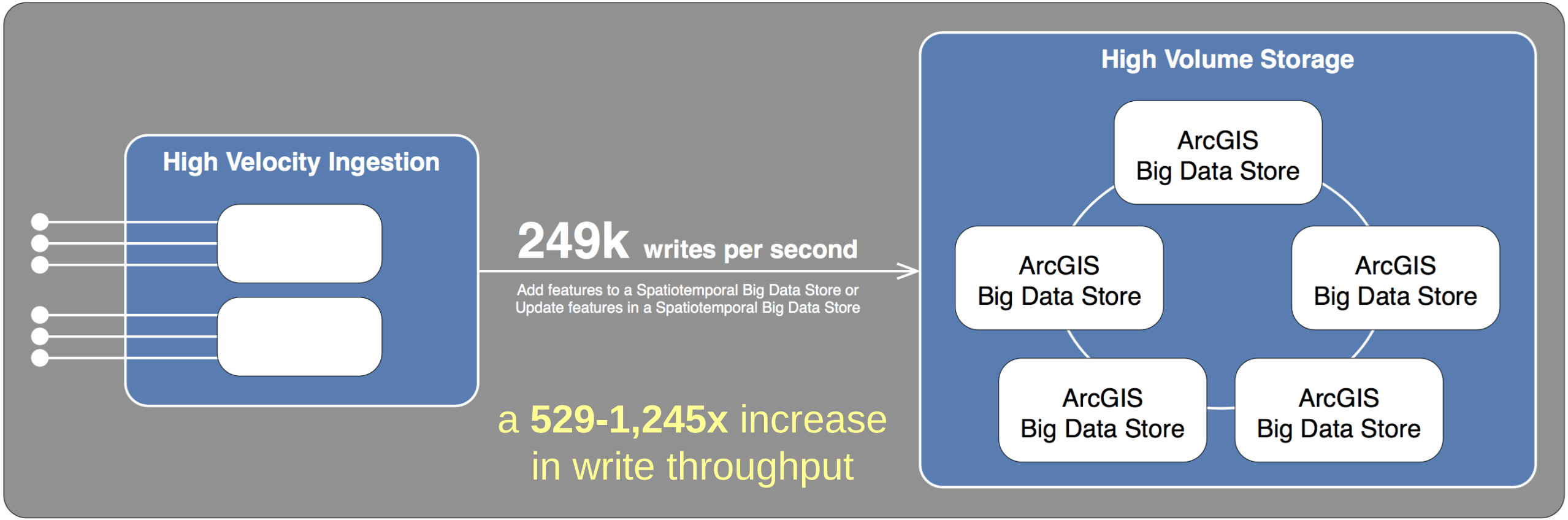
ArcGIS 10.4 Benchmarks *c4.2xlarge (Windows 2012 Server R2): 8 vCPU, 15 GiB, 100GB local SSD, 1,000 Mbps EBS*



Storage Write Throughput	1 node	2 node	3 node	4 node
ArcGIS Big Data Store	106k	143k	192k	224k
ArcGIS Relational Data Store	0.2k	n/a	n/a	n/a

Big Data Store Performance

ArcGIS 10.4 Benchmarks *c4.2xlarge (Windows 2012 Server R2): 8 vCPU, 15 GiB, 100GB local SSD, 1,000 Mbps EBS*



Storage Write Throughput	1 node	2 node	3 node	4 node	5 node
ArcGIS Big Data Store	106k	143k	192k	224k	249k
ArcGIS Relational Data Store	0.2k	n/a	n/a	n/a	n/a

7b. Capacity Planning

Primary Factors to Consider

Big Data Store Performance

What are the primary factors I should consider?

- **Operating environment:** *Amazon EC2 c4.2xlarge*
  - Virtual Machines – beware! resources need to be shared in an effective way, like EC2 or Azure.
  - Bare-Metal machines – dedicated resources and public cloud instances are much more deterministic.
- **Disk**
  - Speed (Mbps) – the faster the better *1,000 Mbps EBS, note: local SSD is much better*
- **Network**
  - Speed (Gbps) – the faster the better. *10 Gbps*
- **RAM**
  - size (GB) – minimum of 16GB is recommended. *15GiB, the Big Data Store was allocated 8GiB*
  - type – minimum of DDR3 is recommended.
  - clock speed (MHz) and transfer rate (Mbps) – the faster the better.
- **Processors**
  - speed (GHz) – the faster the better.
  - # of cores – the more the better. *8 vCPU*

Capacity Planning your Spatiotemporal Big Data Store

Factors to consider

- **Event Velocity**
  - # of events per second
  - Message size (in bytes)
  - Sensor active time per day
  - % of sensors active per day
  - # of peak hours
- **Real-Time node resources:**
  - Real-Time throughput (in e/s)
  - Real-Time additional redundancy (# of nodes)
- **Storage node resources:**
  - Storage total disk size per node (in TBs)
  - Storage retention (in months)
  - Storage replication factor (# of nodes)

7c. Capacity Planning

Estimator Tool

ArcGIS Spatiotemporal Big Data Store Capacity Planning Estimator

note: this estimator only works is targeted at moving observations, if you have stationary, discrete, or change observations please contact Esri to discuss your capacity planning needs.

Parameters:	
Track events per minute when active:	1
Track active time per day (in minutes):	480
Message size (in bytes):	1024
Peak duration (in hours):	4
Peak % of tracks active:	0.8
Non-GPS messages per Track per day:	0
GeoEvent assumed throughput e/s:	2500
GeoEvent desired addt'l redundancy:	1
Storage total disk size per node (in TBs):	3
Storage retention (in # of months):	1
Storage replication factor:	1

Please fill in the green cells with your values.

	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026
Expected # of unique Track IDs:	100000	200000	300000	400000	500000	600000	700000	800000	900000	1000000	1100000

Calculations:	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026
Events per day:	38,400,000	76,800,000	115,200,000	153,600,000	192,000,000	230,400,000	268,800,000	307,200,000	345,600,000	384,000,000	422,400,000
Events per second on avg:	444	889	1,333	1,778	2,222	2,667	3,111	3,556	4,000	4,444	4,889
Events per second during peak:	2,133	4,267	6,400	8,533	10,667	12,800	14,933	17,067	19,200	21,333	23,467

Node Estimates:											
# of GeoEvent nodes:	2	3	4	5	6	7	7	8	9	10	11
# of Big Data Store nodes:	2	2	3	3	3	4	4	5	5	5	6
Total nodes:	4	5	7	8	9	11	11	13	14	15	17

note: to determine your throughput you should benchmark with illustrative real-time analytics representing your planned usage.

note: GeoEvent nodes = (events per second during peak / throughput) + addt'l redundancy, note: assumes linear scalability and an 8 core, 32GB RAM machines

*note: Storage nodes = ((events per day * message size * 30 days * retention months) / storage size) + additional redundancy, note: assumes no compression*

note: Storage volume capacity needed is completely dependent on the retention requirements of the data, the longer the retention the more nodes that will be needed

10.4

Summary

Archiving & Visualizing Observation Data

- **ArcGIS 10.4 has been enhanced so you:**
  - **Can ingest real-time HIGHER velocity data and CAN store it.**
  - **Can visualize high velocity & volume data in feature form, in AGGREGATE form, LIVE & HISTORICALLY.**
  - **Visualization CAN work with high volume data.**

- **To learn more:**
  - **See the tutorial:**
    'ArcGIS Spatiotemporal Big Data Store'

Real-Time GIS

Other session

- **Real-Time GIS: Applying Real-Time Analytics**
  Thu Apr 28, 8:30 – 10:00, Grand Ballroom B

Thank you!

Mark Bramer

mbramer@esri.com