



DIRECTIONS

LIVE Online



DIRECTIONS

LIVE Online

Big Data Analysis in ArcGIS

ArcGIS GeoAnalytics Engine





Fepri Putra Panghurian

Analytic and Data Management Team
lead
Esri Indonesia

Big Data Processing

Big Data Characteristics

Volume

- Large data would not fit on a single machine.
- Specialized tools and frameworks are required to store process and analyze it.

Velocity

- How fast the data is generated.
- Specialized tools are required to ingest such high velocity data into the big data infrastructure and analyze the data in real-time.

4'v

Variety

- The forms of the data.
- Consists of structured, unstructured, or semi-structured data, including text data, image, audio, video and sensor data.

Veracity

- How accurate is the data.
- Cleansing of data is important so that incorrect and faulty data can be filtered out.

Big Data Tools

Data Storage



Ingestion



Arcgis
Gevent Server



Analysis



ArcGIS
Geoanalytics

Visualization

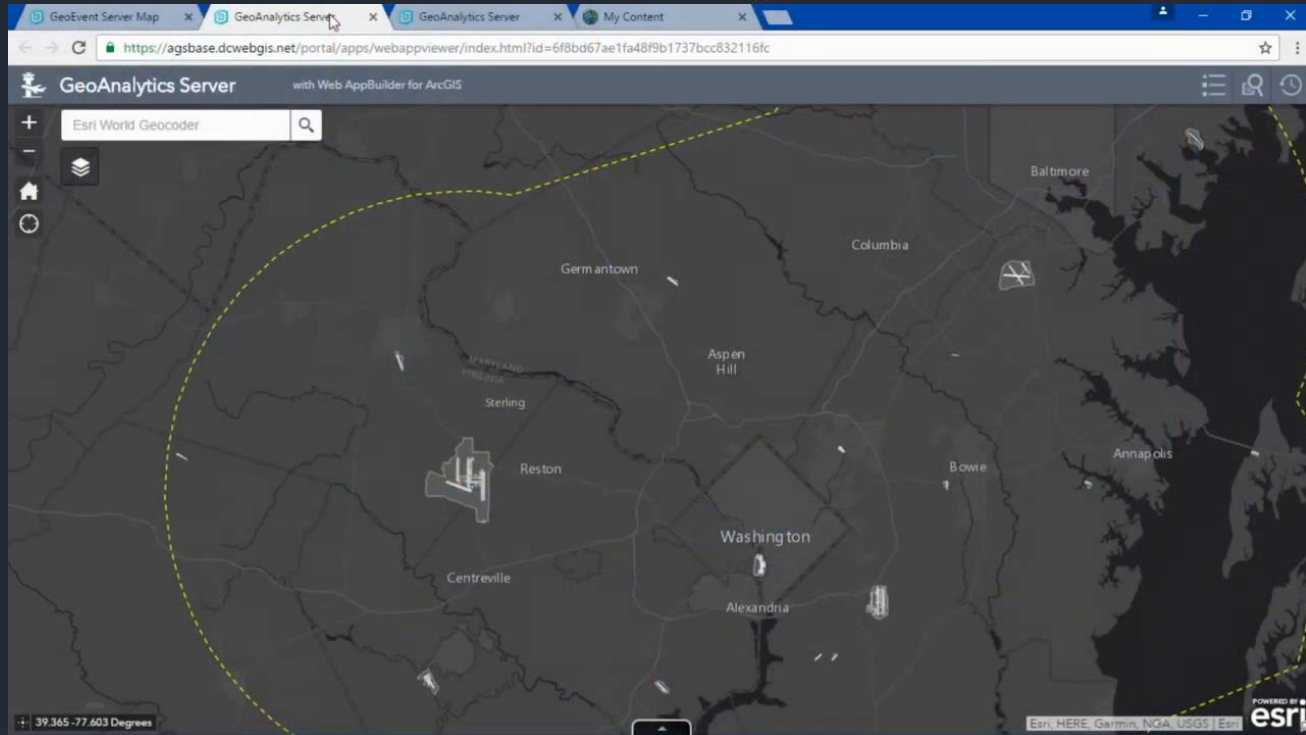


Benefit Spatial Big Data Analysis

(Demo 1 – Search Cluster From New York Uber Pickup Dataset)



Benefit Spatial Big Data Analysis (Demo 2 – Browse Large Dataset)



Benefit Spatial Big Data Analysis (Demo 3 – Detect Incident)

Ground Level Ozone

Causes difficulty breathing and inflammation of the airways.

Leads to:

- Increased medication use
- Emergency room visits
- Time off work
- Aggravated asthma symptoms
- Many more health concerns

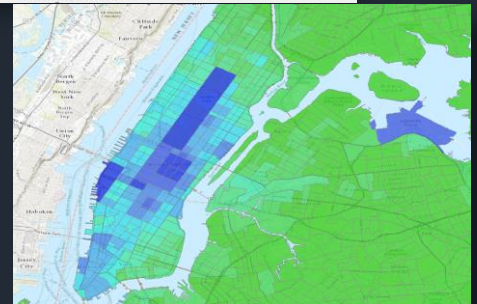
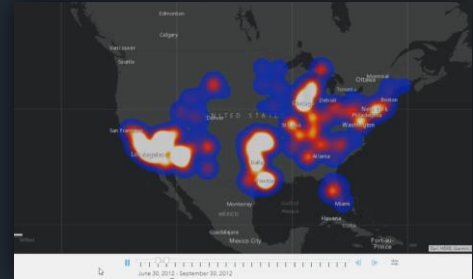
Arcgis GeoAnalytics

GeoAnalytics

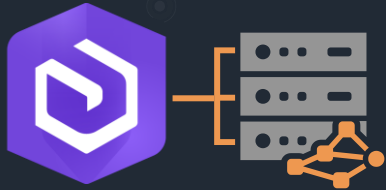
GeoAnalytics *parallelizes computing* to quickly analyze large amounts of *vector and tabular data*



A collection of analysis tools to identify *patterns, relationships, anomalies* and *incidents* in large amounts of data across space and time



GeoAnalytics Product Catalog



GeoAnalytics Server
Distributed processing across
multiple *server cores and*
machines with ArcGIS

Enterprise

**Requirements: Enterprise +
GeoAnalytics Server License**



GeoAnalytics Desktop
Parallel processing across *cores*
on your laptop or desktop with
ArcGIS Pro

**Requirements: Advanced
License**



GeoAnalytics Engine
Parallel processing across *cores*
using your current apache
spark cluster

Analysis Capabilities

Summarize Data

- Aggregate Points
- Build Multi-Variable Grid
- Describe Dataset
- Join Features
- Reconstruct Tracks
- Summarize Attributes
- Summarize Within

Use Proximity

- Create Buffers

Analyze Patterns

- Calculate Density
- Create Space Time Cube
- Find Hot Spots
- Find Point Clusters *+
- Forest-based Classification and Regression
- Generalized Linear Regression

Data Enrichment

- Enrich from Multi-Variable Grid

Find Locations

- Detect Incidents
- Find Similar Locations
- Find Dwell Locations +
- Geocode Locations

Manage Data

- Append Data
- Calculate Field
- Clip Layer
- Copy to Data Store
- Dissolve Boundaries
- Merge Layers
- Overlay Layers

What types of data can I analyze? And from where?



- Access and share data within Enterprise with ***your Enterprise portal***



- Seamlessly analyze data collected with ***ArcGIS GeoEvent Server***



- Analyze data in ***Hive, HDFS, and files***



- Connect to Azure and Amazon ***cloud stores***

Easily Connect To Your Big Data

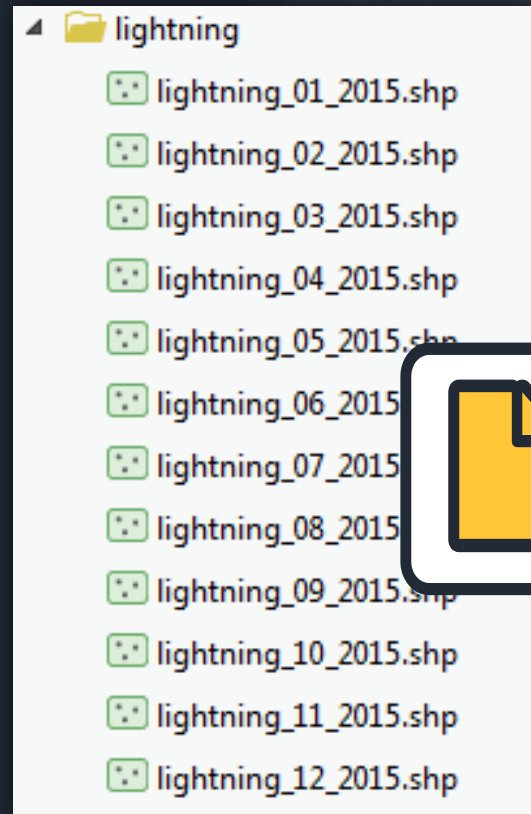
Big Data File Shares

Read directly from files stored in

- Hive
- HDFS
- Shared folders
- Cloud stores

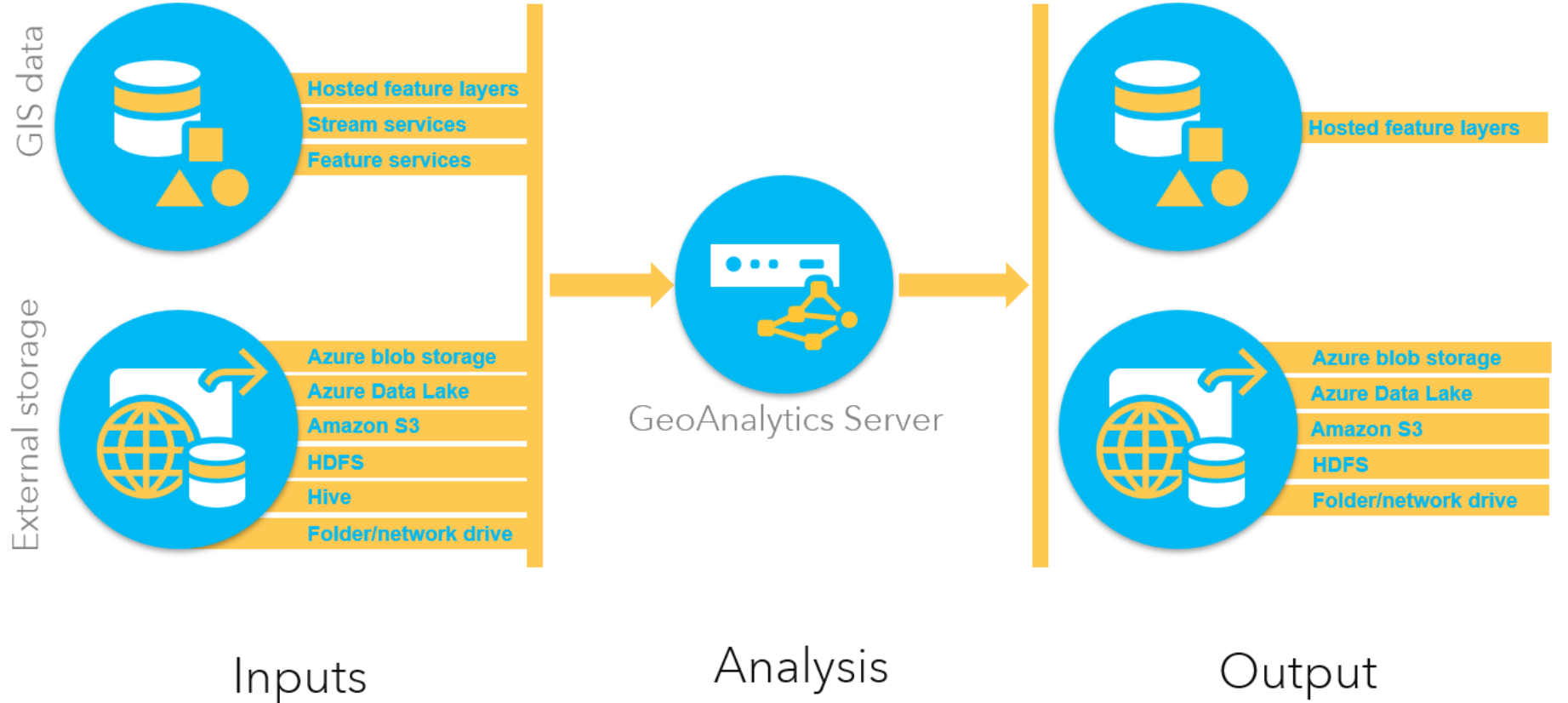
Supported file types include

- Delimited files
- Shapefiles
- ORC
- Parquet

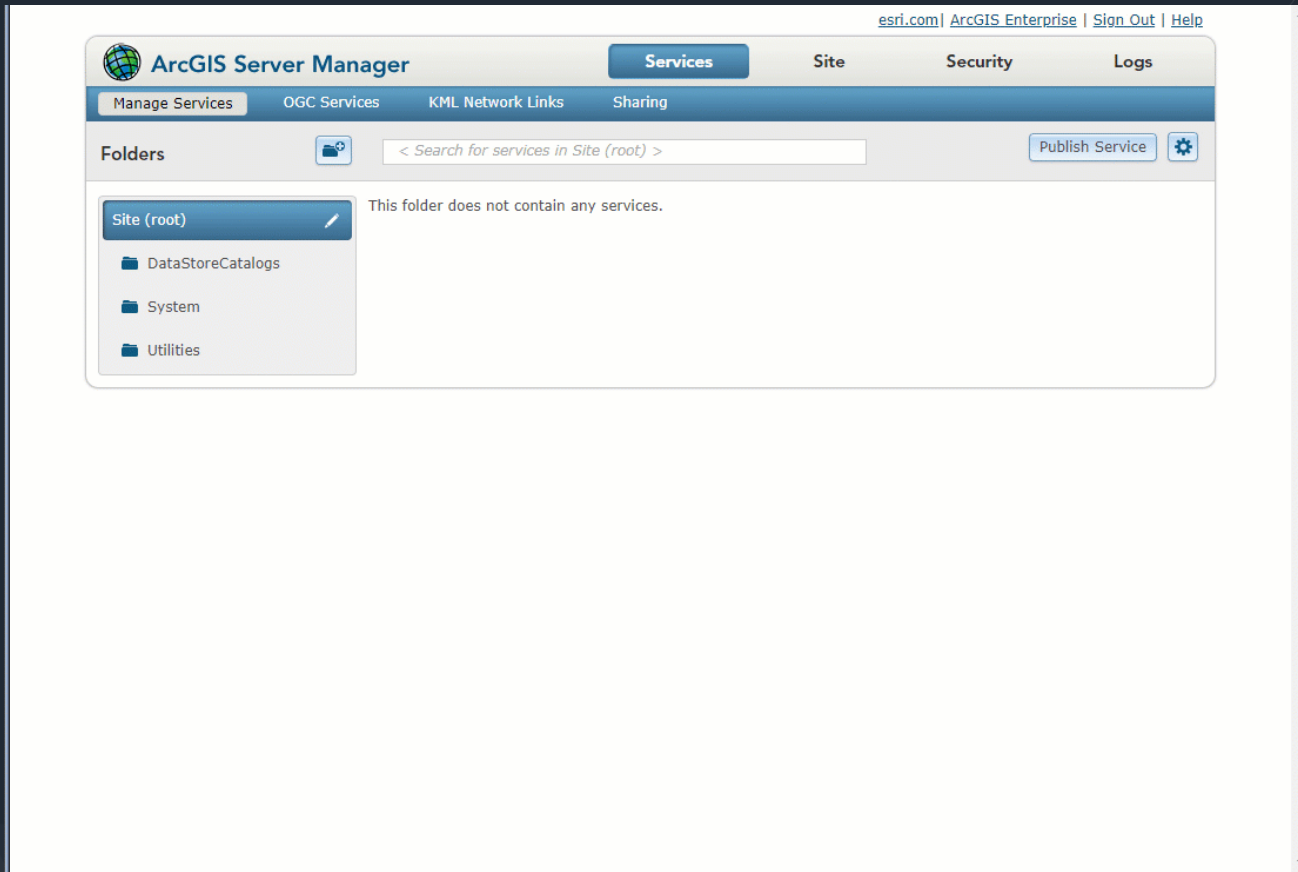


Use multiple files with a matching schema as a single dataset

Write to big data file shares



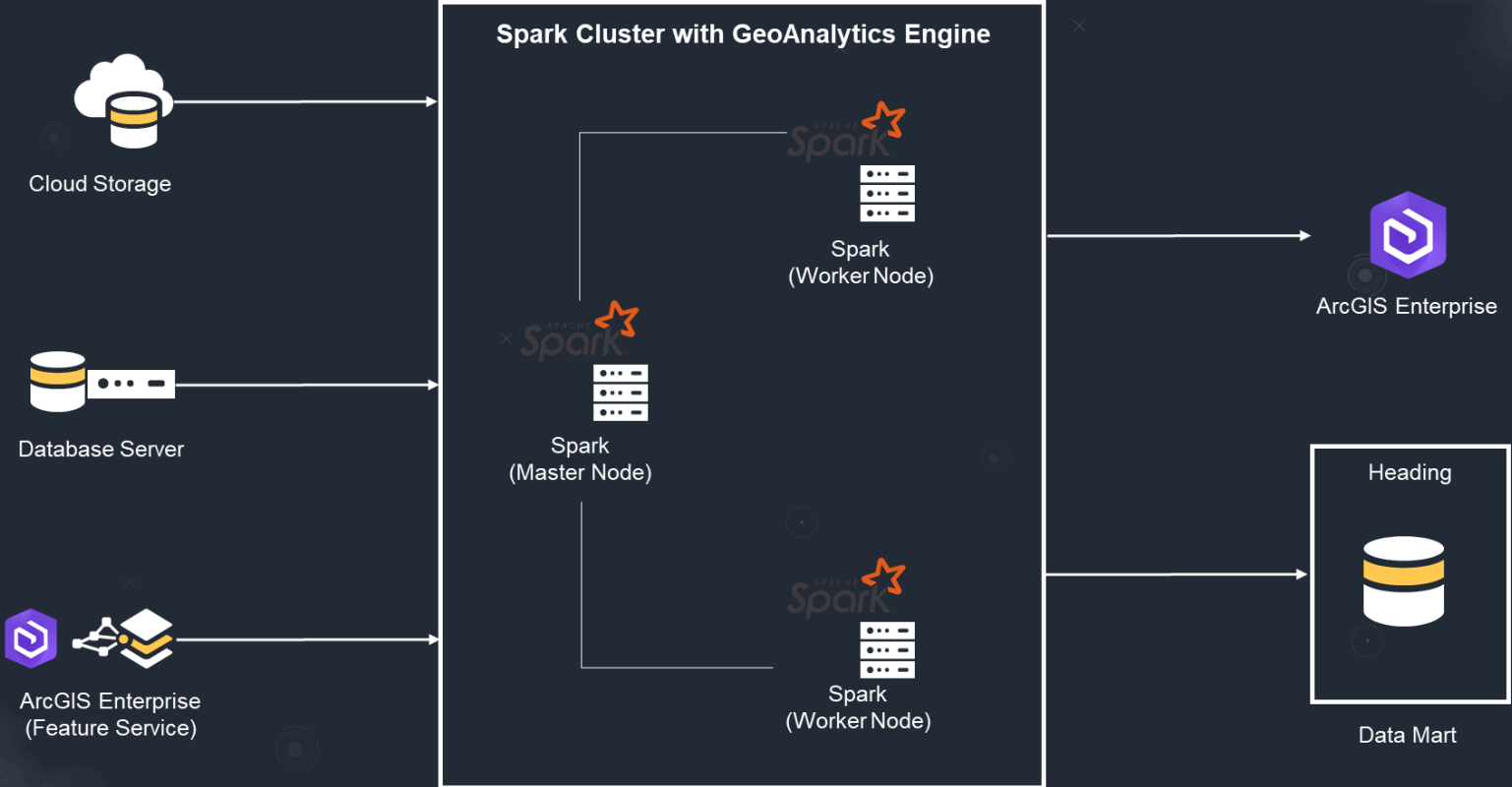
How do I actually use all those data sources?



The screenshot displays the ArcGIS Server Manager web interface. At the top right, there are links for [esri.com](#), [ArcGIS Enterprise](#), [Sign Out](#), and [Help](#). The main navigation bar includes **ArcGIS Server Manager** and tabs for **Services**, **Site**, **Security**, and **Logs**. Below this, a secondary navigation bar contains **Manage Services**, **OGC Services**, **KML Network Links**, and **Sharing**. The **Services** tab is active, showing a **Folders** section with a search bar containing the text "< Search for services in Site (root) >". To the right of the search bar is a **Publish Service** button and a settings gear icon. The folder list on the left includes **Site (root)** (selected), **DataStoreCatalogs**, **System**, and **Utilities**. The main content area for the selected folder displays the message: "This folder does not contain any services."

Arcgis GeoAnalytics Engine

GeoAnalytics Engine Architecture



GeoAnalytics Engine Licensing

The screenshot shows the Esri GeoAnalytics Engine subscription page. The page is titled "Esi Indonesia Big Data Team" and features the Esri logo. Key information includes:

- Subscription ID: 639037818
- License Type: ArcGIS GeoAnalytics Engine
- Plan: Prepaid
- Learn more: [Learn more](#)
- Created: Nov 1, 2022
- Renewal date: Oct 31, 2023

Core hours usage is displayed in two sections:

- Core hours @:** Total remaining core hours: 99,999.75. Last 30 days: 0.25. Last 24 hours: 0.25.
- Core hours:** Last 12 months: 0.25. Last 6 months: 0.25.

Both sections include line graphs showing core hours consumed over time.

```
Untitled1.ipynb | Untitled1.ipynb | hurricanes_data.ipynb | Python 3 (pykernel)

[3]: # Log in
import geoanalytics
geoanalytics.auth(username="esriid_bigdata", password="Esr1123412022")

# Imports
from geoanalytics.tools import ReconstructTracks
from geoanalytics.tools import SpatiotemporalJoin
from geoanalytics.sql import functions as ST

# Path to the hurricane track points data and USA counties data
hurricanes_data_path = r"https://services2.arcgis.com/FiaPA4ga0lQkduv3/arcgis/rest/" \
    "services/IBTrACS_ALL_list_v04r00_points_1/FeatureServer/0"
counties_data_path = "https://services.arcgis.com/P3ePLHYs2RvCk3/ArcGIS/rest/services/" \
    "USA_Counties_Generalized/FeatureServer/0"

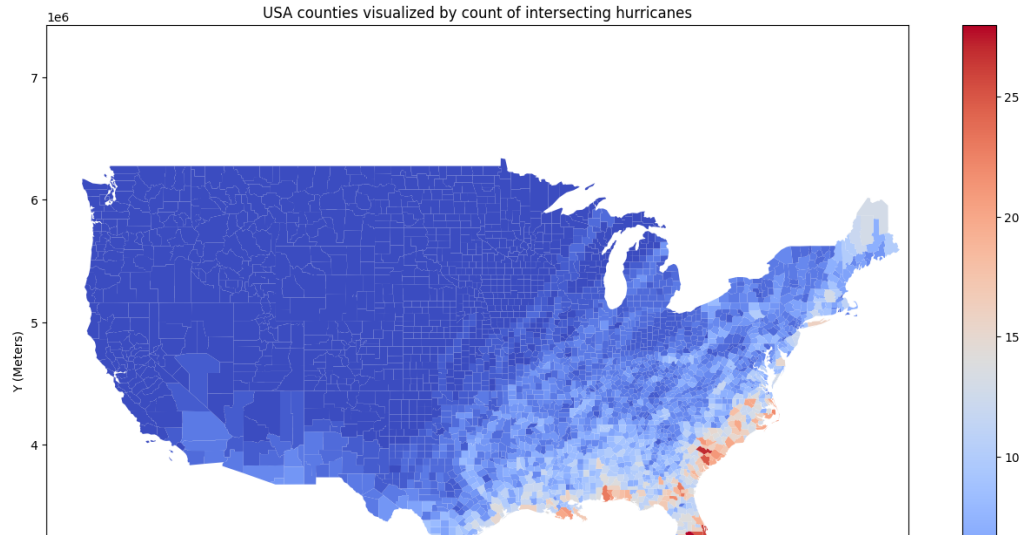
# Create DataFrames for hurricane track points and USA counties
hurricanes_df = spark.read.format("feature-service").load(hurricanes_data_path)
hurricanes_df = hurricanes_df.withColumn("usa",
    ST.env_intersects("shape",
        xmin=-14165303, xmax=-7384568,
        ymin=1647768, ymax=7426938)).where("usa=true")
counties_df = spark.read.format("feature-service").load(counties_data_path)

# Reconstruct hurricane points into tracks using ReconstructTracks
rt_result = ReconstructTracks() \
    .setTrackFields("NAME") \
    .setDistanceMethod(distance_method="Geodesic") \
    .run(dataframe=hurricanes_df)

# Use Spatiotemporal Join to join the hurricane tracks to each county
join_result = SpatiotemporalJoin() \
    .setLeftJoin(left_join=True) \
    .setJoinOnTime() \
    .setSpatialRelationship(spatial_relationship="Intersects") \
    .run(target_dataframe=counties_df, join_dataframe=rt_result)
```

GeoAnalytics Engine (Demo 1)

```
[4]: # Plot the join result and visualize the frequency of hurricanes for each county
result_plot = join_result.st.plot(cmap_values="COUNT",
                                cmap="coolwarm",
                                legend=True,
                                figsize=(16,10),
                                geometry="shape")
result_plot.set_title("USA counties visualized by count of intersecting hurricanes")
result_plot.set_xlabel("X (Meters)")
result_plot.set_ylabel("Y (Meters)")
result_plot.set_xlim(left=-14165300, right=-7384568)
result_plot.set_ylim(bottom=1947768, top=7426938);
```



Benefit Using ArcGIS In Big Data

- ✓ Use Spatial Analyst (Using ArcGIS) Alongside with current Big Data Stack
- ✓ Perform fast because using Spark Server as parallel processing not only in Desktop (GeoAnalytics Server and GeoAnalytics Engine)
- ✓ Could integrate with another big data system and storage (Kafka, Cloud Storage, Datalake, etc)
- ✓ Cost Effective (Pay as computed hours) and there is not core limitation (like GeoAnalytics Server)
- ✓ Many spatial analyst could be handled in Big Data processing using Arcgis GeoAnalytics.



Fepri Putra Panghurian

Analytic and Data Management Team
lead
Esri Indonesia



Hubungi kami

UP NEXT:



DIRECTIONS

LIVE Online

**Industry
focus**





DIRECTIONS

LIVE Online