DIRECTIONS

LIVE Online

esri Indonesia

DIRECTIONS

LIVE Online

Big Data Analysis in ArcGIS

ArcGIS GeoAnalytics Engine

esri Indonesia

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

Geoevent Server

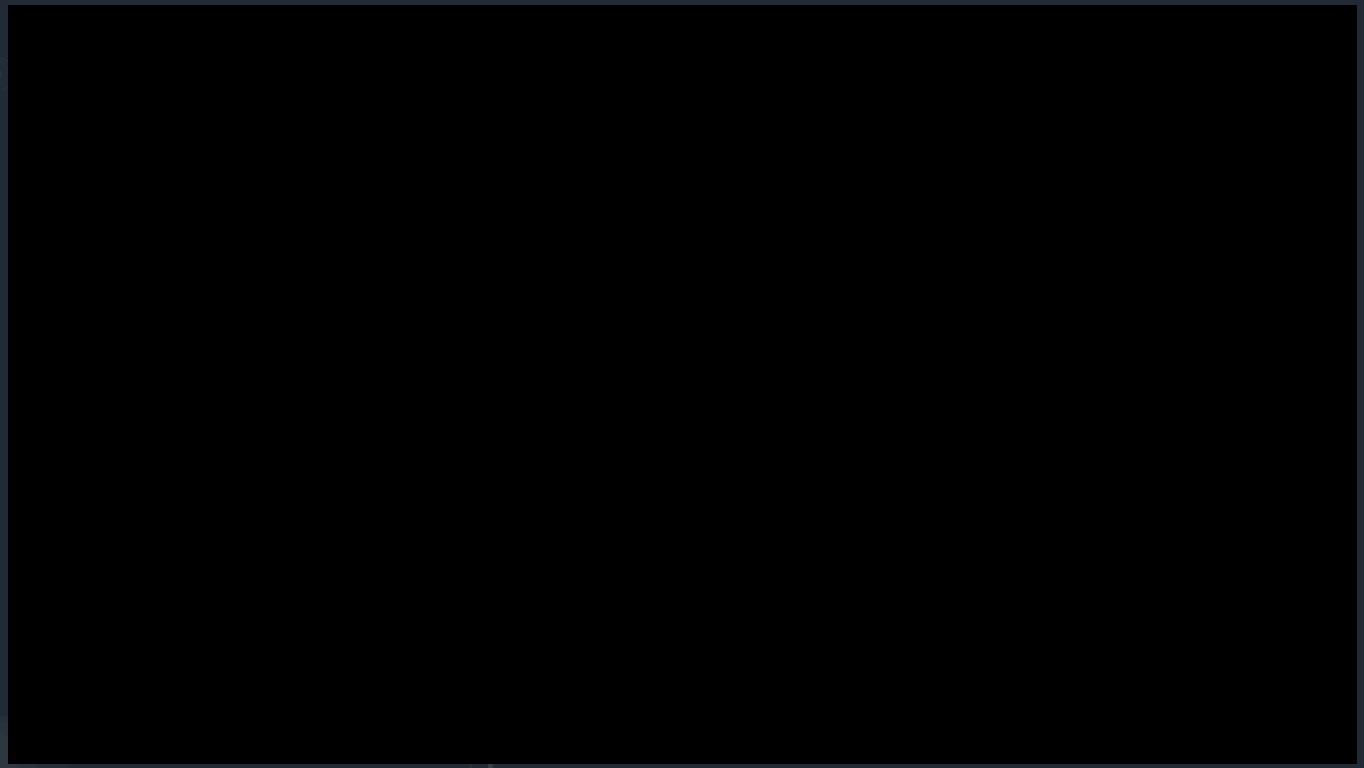
Analysis

ArcGIS

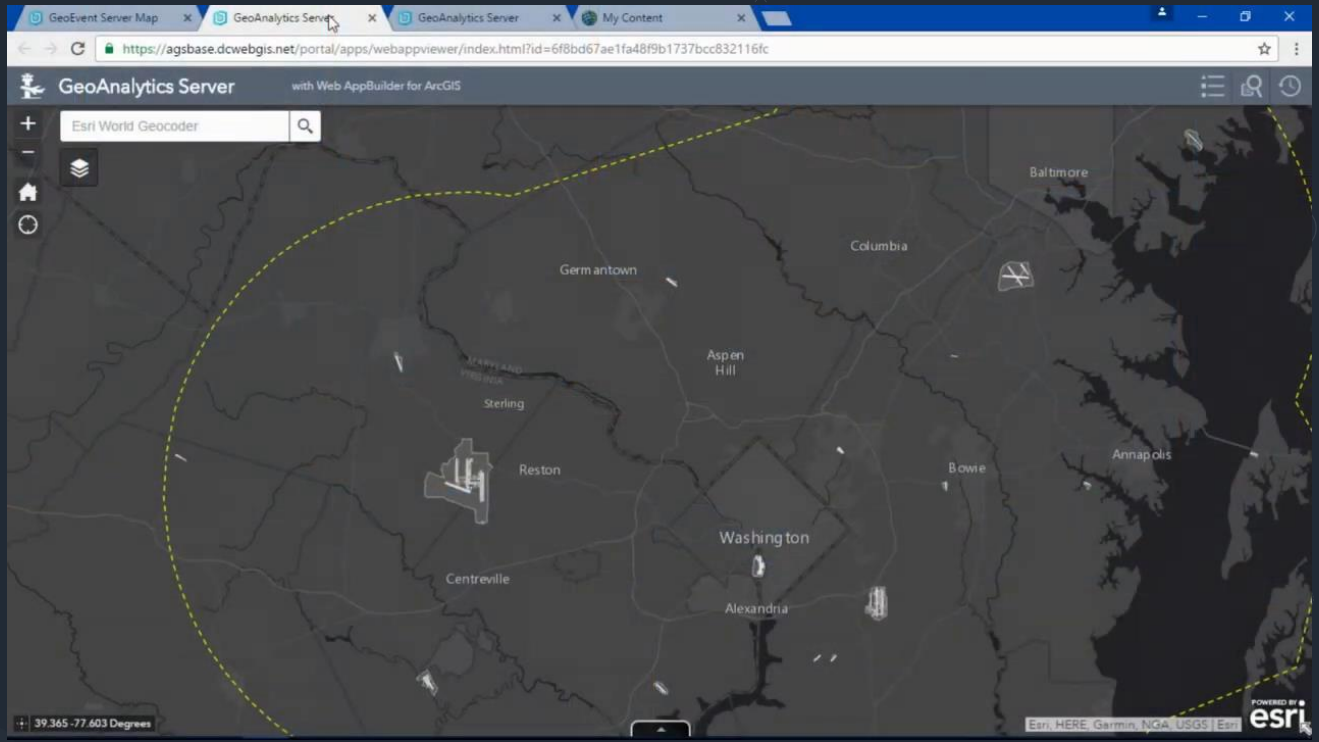
Geoanalytics

Visualization

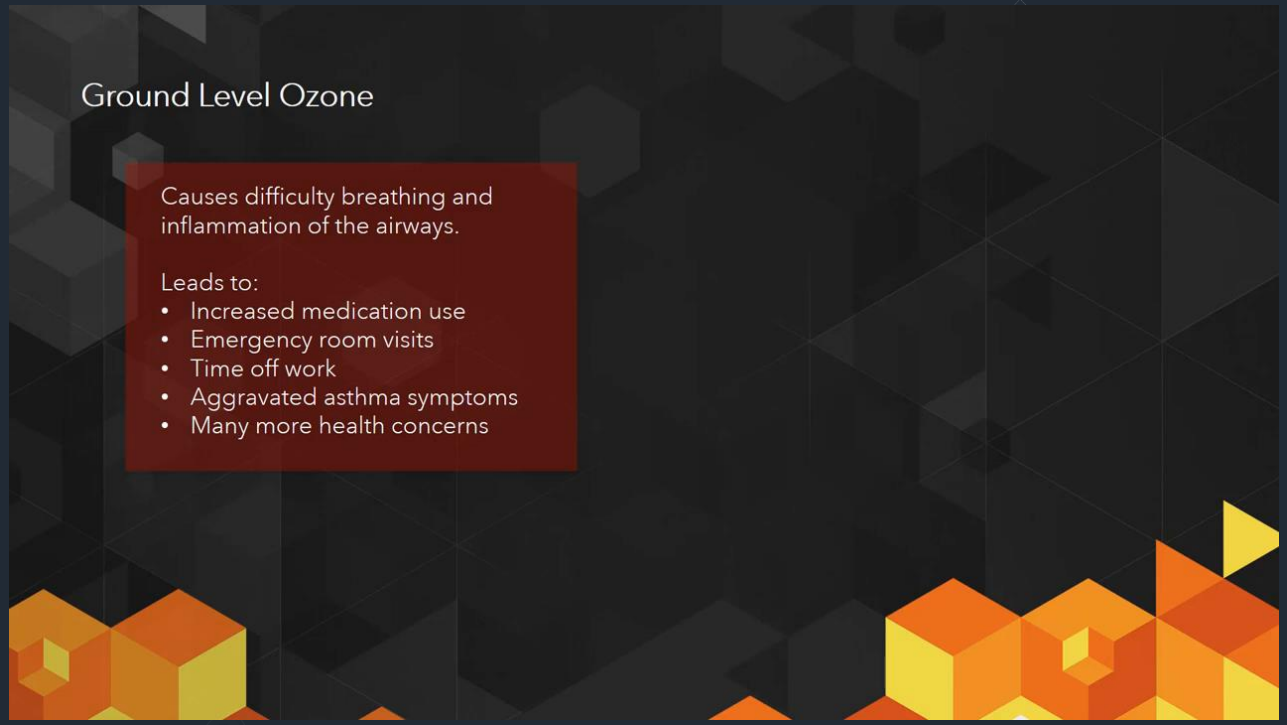
Benefit Spatial Big Data Analysis
(Demo 1 – Search Cluster From New York Uber Pickup Dataset)



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



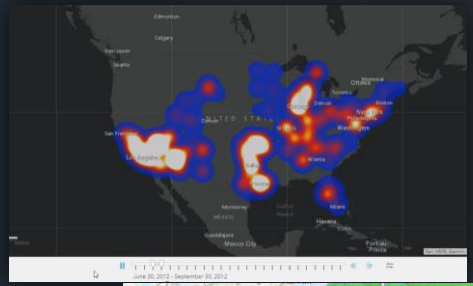
Benefit Spatial Big Data Analysis (Demo 3 – Detect Incident)



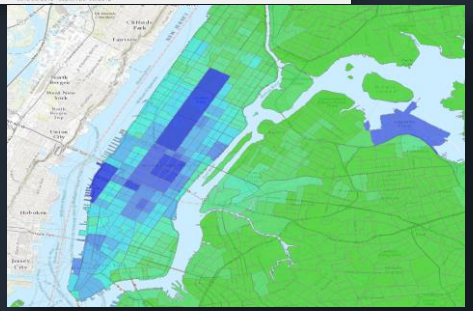
Arcgis GeoAnalytics

GeoAnalytics

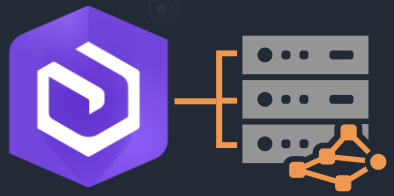
GeoAnalytics *parallelizes computing* to quickly analyze large amounts of *vector and tabular data*



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

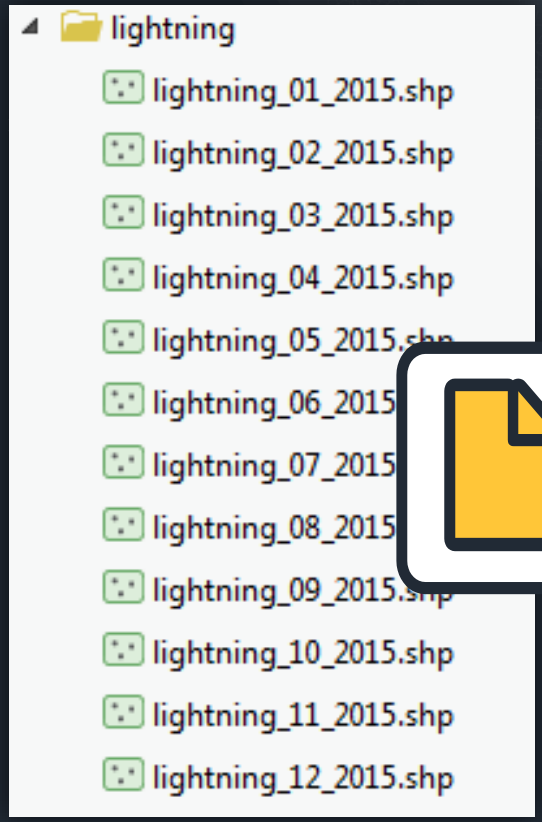
Big Data File Shares

Read directly from files stored in

- **Hive**
- **HDFS**
- **Shared folders**
- **Cloud stores**

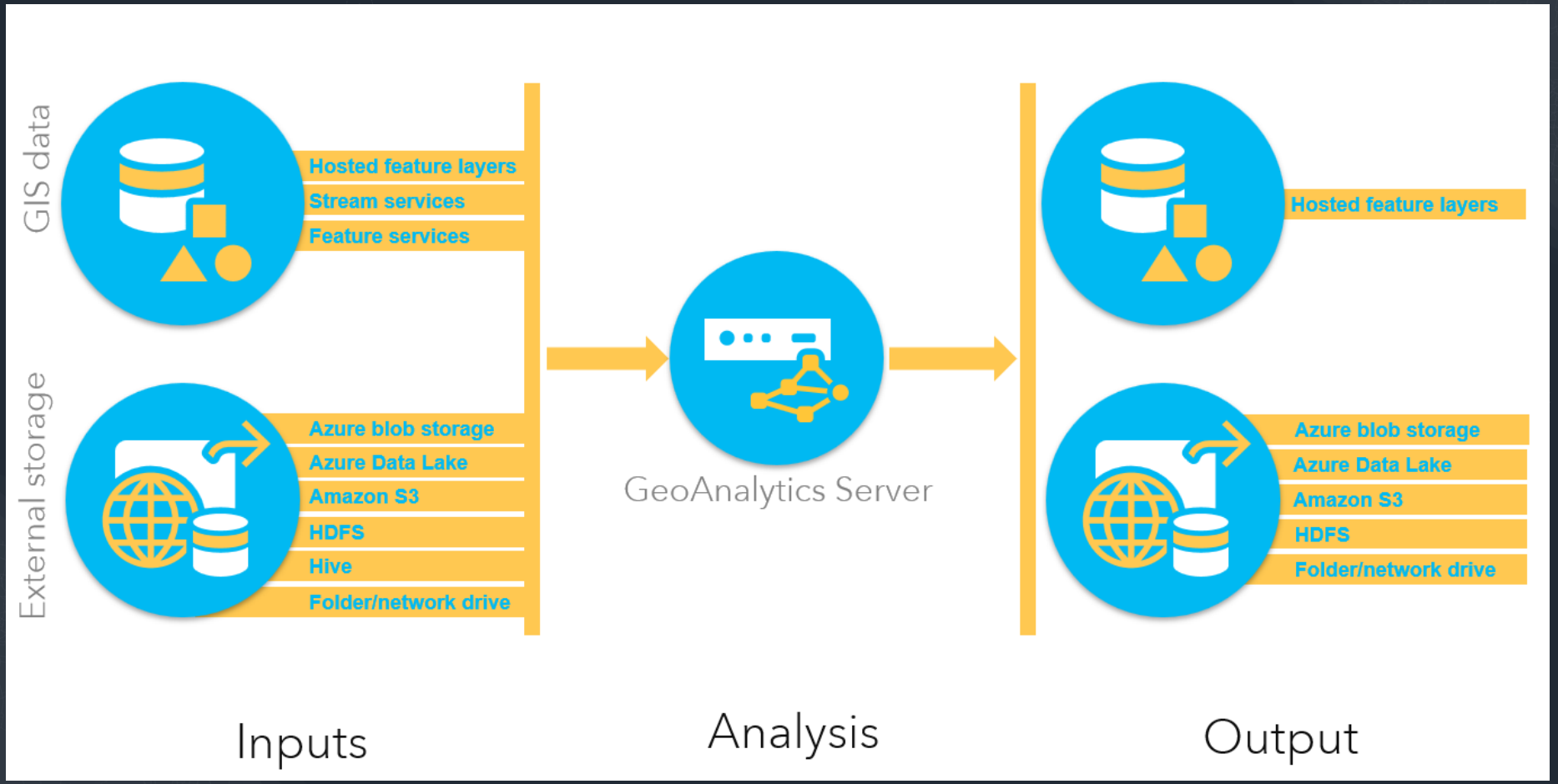
Supported file types include

- **Delimited files**
- **Shapefiles**
- **ORC**
- **Parquet**

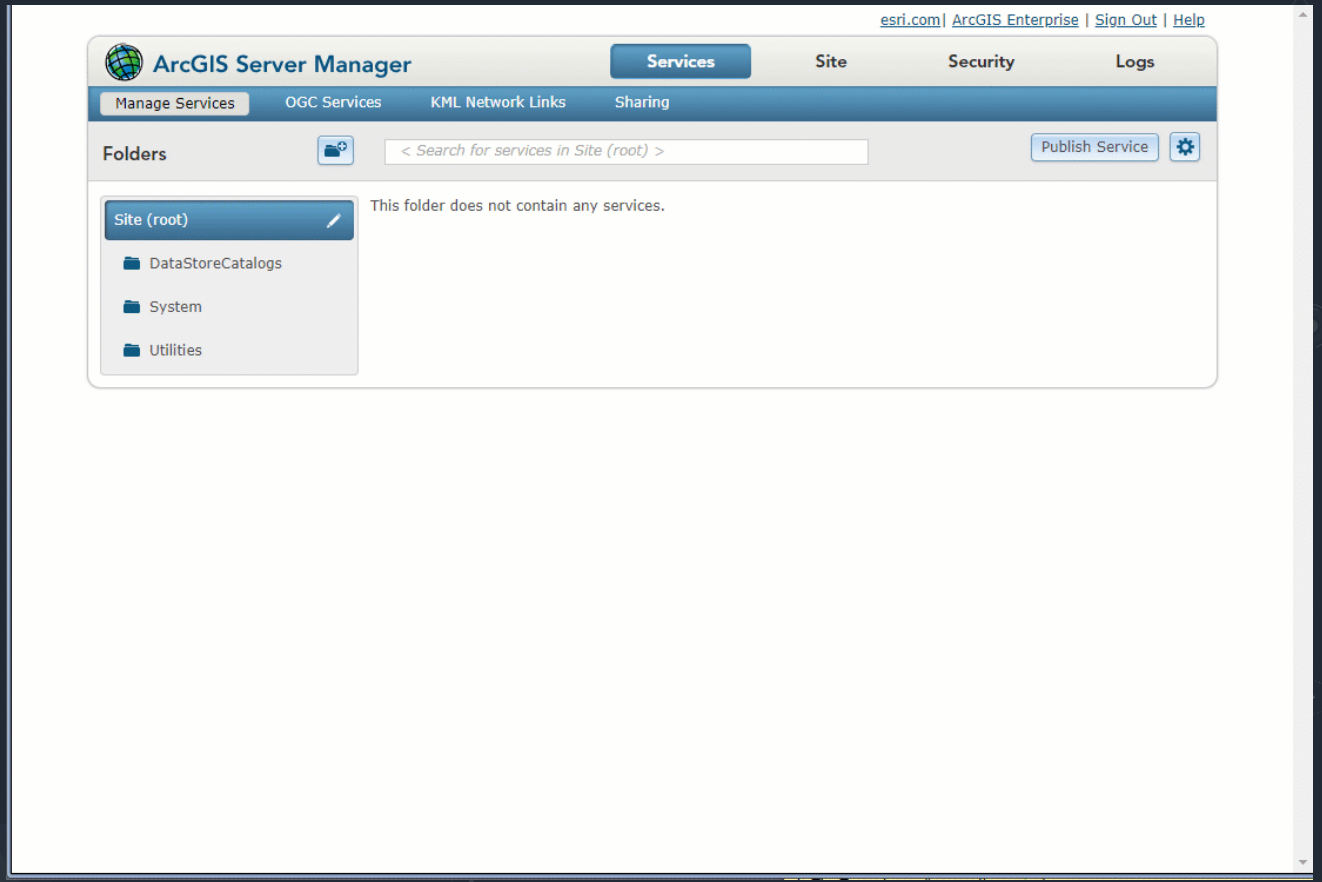


Use multiple files with a matching schema as a single dataset

Write to big data file shares

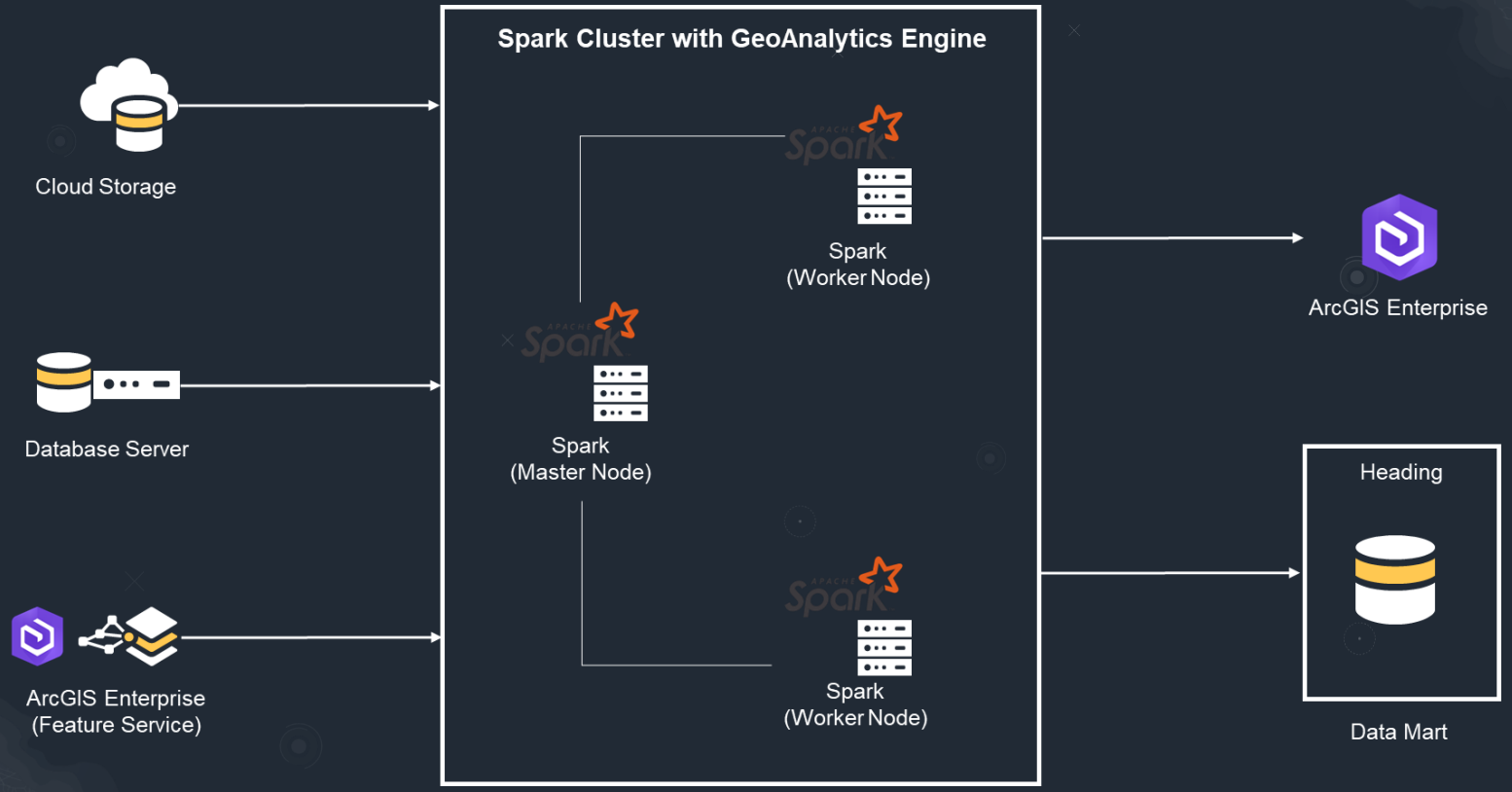


How do I actually use all those data sources?

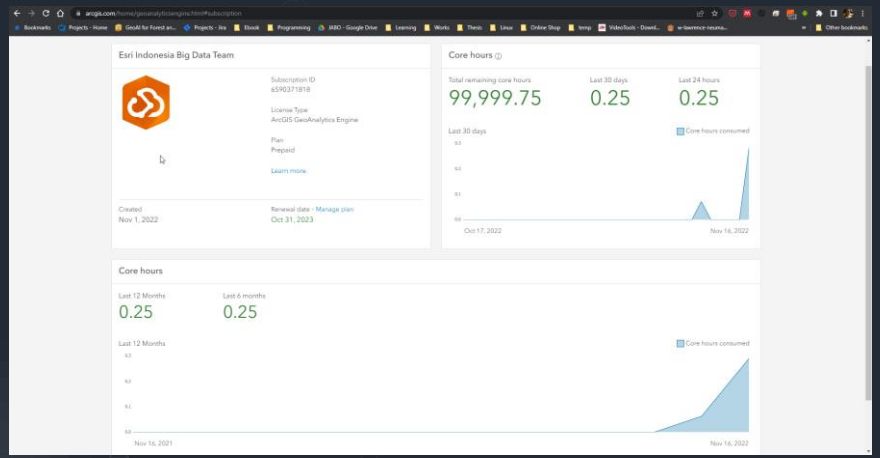


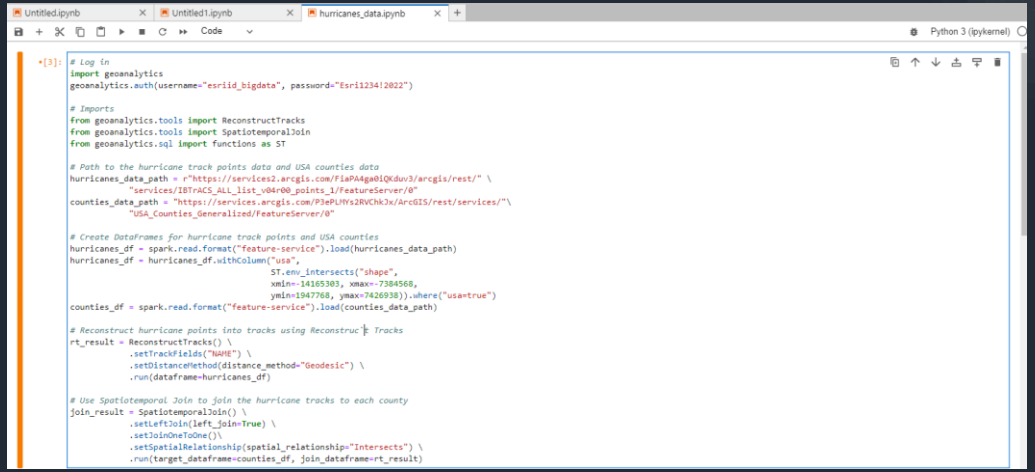
Arcgis GeoAnalytics Engine

GeoAnalytics Engine Architecture



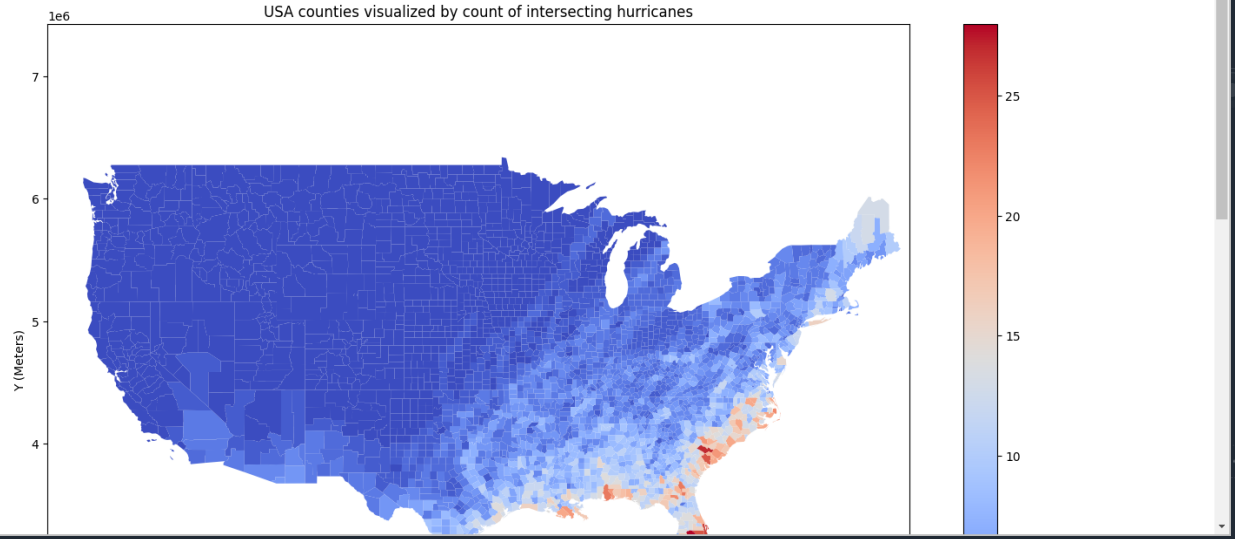
GeoAnalytics Engine Licensing





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=-14165303, 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 (GeoAnalyutics 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

DIRECTIONS LIVE Online

UP NEXT:

DIRECTIONS

LIVE Online

esri Indonesia