### Intra-agency Data Commons (data mesh node)

Server Components (Docker composed together)

### Deployment

## Terraform script

- Allocates cloud compute, storage, security groups.

### Domain Initialization

- Via Dockerfile RUN
- commands - Initial setup/ingest of static resources
- Optionally regester this Data Commons instance with catalog server
- Or manually add other external Data Commons for data discovery

# Intra-agency STAC Catalog

- implemented via pgSTAC/PostGIS

## STAC API IIIII

- Real-time, dynamic querying, filtering, and retrieval of data
- provides RESTful interface to intra-agency STAC catalog
- implemented via Python/FastAPI

#### STAC Browser GUI

- Browser-based data discovery tool
- Search for data products both internal and external to local STAC catalog
  - Kick off new processing
- workflows with search results

# JupyterHub 🌥

- Can receive discovered search results as input with blank workspace
- Example Jupyter notebooks provided
- 3DEP libraries preinstalled (Entwine, PDAL, GDAL)
- long-term: Databricks, or Pangeo Cloud Federation JupyterHub Kubernetes

### GSA PoC Customization of STAC API: RegisterDataProduct()

### Baseline API endpoint definition:

- STAC API Extension Endpoint 'Transaction'
- Supports HTTP methods
- POST, PUT, PATCH, DELETE - Input body: 'STAC Item' GeoJSON Feature augmented with 'Foreign

Members'

## GSA Customization 1: Self-reported ID OAuth2/IAM/AD/ID.me (Stub for PoC)

### GSA Customization 2: Provenance IIIII

- Enforce existence of 'Link' object of relation type='derived\_from'
- could be upstream data service endpoint URL, filename, REST API call, GUI-based map service, etc

### GSA Customization 3: Data Product Info III

- Jupyter notebook filename/URL (register new notebook now?)
- Source code has to be referenceable
- Plain language description - return value type of data
- product (Stub for PoC)

### Legend

Four Pillars of Data Mesh: Domain-Oriented Decentralized Data Ownership

- IIIII Data as a Product
- 🌥 Self-service Data Infrastructure as a Platform - 🖪 Federated

Computational Governance

Hover over/click boxes for

links to definitions.