

INFRAESTRUCTURA COLOMBIANA DE DATOS ESPACIALES (ICDE)

EAFIT

- 1 Andrés Guerra
- 2 Alejandro Mc Ewen
- 3 Tomás Calle
- 4 Juan Felipe Ortiz
- 5 Tomás Duque

PROBLEMÁTICA

1. El ICDE recopila y procesa cientos de datos geográficos del país, pero carece de herramientas robustas para validarlos.
2. El ICDE busca validar datos nuevos y existentes, además de generar estadísticas sobre la calidad de estos datos.

DEFINICIONES IMPORTANTES

Metadato: Información que describe o da información importante sobre o alrededor de un dato.

Dato: Unidad de información o servicio geográfico.

METADATOS -REQUISITOS



Campo	Vector	Raster	Geoservicio
Código	✓	✓	✓
Idioma	✓	✓	✓
Codificación de caracteres	✓	✓	✓
Contacto de los metadatos	✓	✓	✓
- Rol	✓	✓	✓
- Nombre de la Organización	✓	✓	✓
- Nombre del cargo	✓	✓	✓
- Número de teléfono	✓	✓	✓
- Tipo de número de teléfono	✓	✓	✓
- Dirección	✓	✓	✓
- Ciudad	✓	✓	✓
- País	✓	✓	✓
- Enlace	✓	✓	✓
Fecha del metadato	✓	✓	✓
- Tipo de fecha	✓	✓	✓
Norma de metadatos	✓	✓	✓
Perfil de metadatos	✓	✓	✓
Ámbito de los metadatos	✓	✓	✓

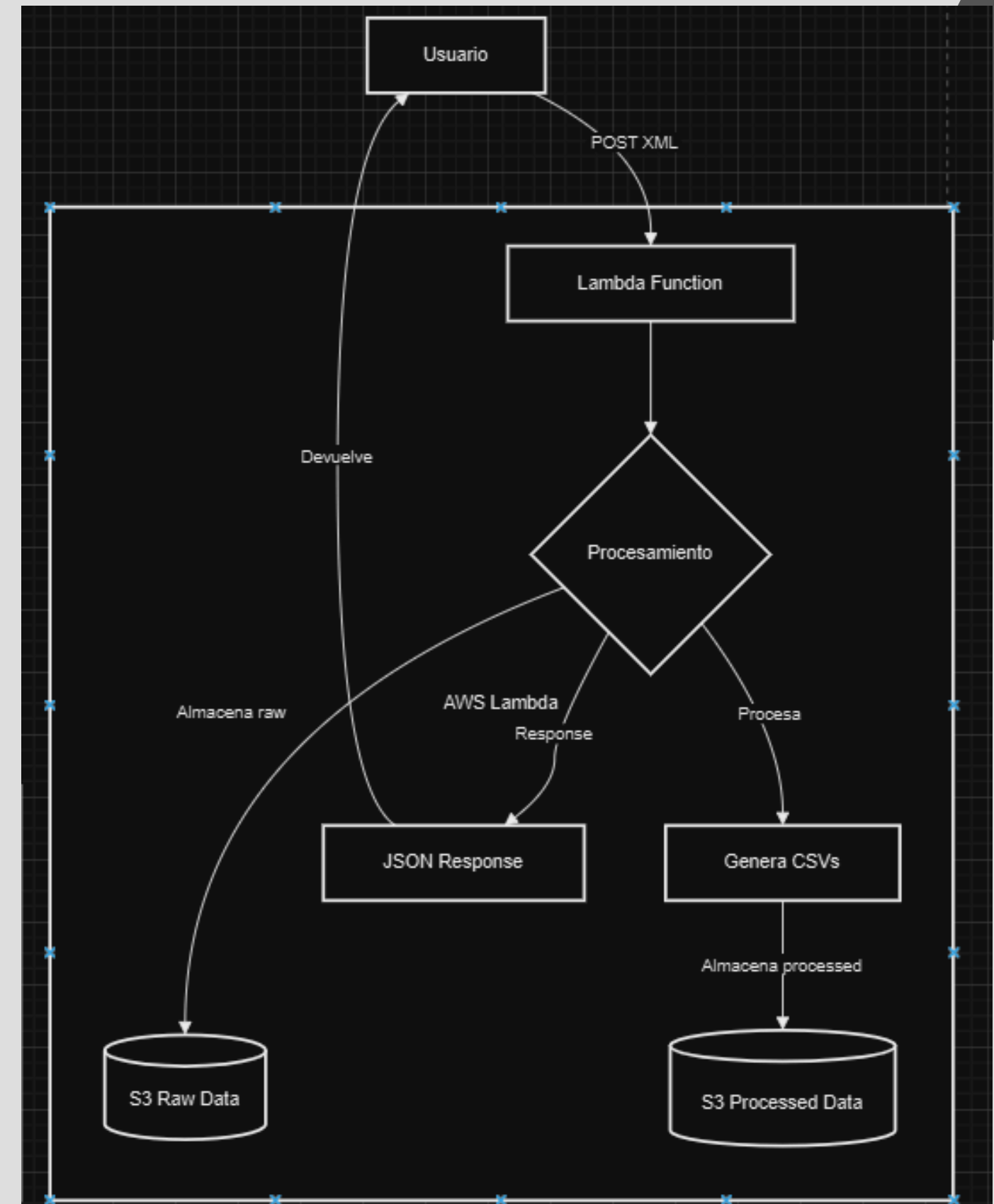
Campo	Vector	Raster	Geoservicio
Citación	✓	✓	✓
- Título	✓	✓	✓
- Fecha	✓	✓	✓
- Tipo de fecha	✓	✓	✓
Resumen	✓	✓	✓
Contacto	✓	✓	✓
Tipo de representación espacial	✓	✓	✓
Resolución espacial	✓	-	-
Categoría del tema	✓	✓	✓
Extensión geográfica	✓	✓	✓
Mantenimiento del recurso	✓	✓	✓
Palabras clave descriptivas	✓	✓	✓

METADATOS -REQUISITOS

Campo	Vector	Raster	Geoservicio
Información de calidad (Linaje)	✓	✓	✓
Sistema de referencia	✓	✓	✓
Información de distribución	✓	✓	✓
Identificación del servicio	-	-	✓
Información de operación	-	-	✓

CICLO DE VALIDACIÓN, PROCESAMIENTO

Metadatos para
Geoservicio, Vector, Raster



METADATOS - VALIDACIÓN

Vector, Raster

The screenshot displays a REST client interface with a dark theme. At the top, a POST request is configured to the URL `https://wphf4ufcrf.execute-api.us-east-1.amazonaws.com/prod/upload-metadata`. The 'Body' tab is selected, showing a file named `test5.xml` uploaded from the local system. The request was sent successfully, returning a `200 OK` status. The response body is shown in the 'Body' tab, formatted as JSON. The JSON structure includes a `metadata` object with a `metadato` object containing various fields like `codigo`, `idioma`, `codificacion`, `contacto`, `direccion`, `ciudad`, `pais`, and `email`.

Request Details:

- Method: POST
- URL: `https://wphf4ufcrf.execute-api.us-east-1.amazonaws.com/prod/upload-metadata`
- Body Type: binary
- File: `test5.xml`

Response Details:

- Status: 200 OK
- Time: 4.84 s
- Size: 3.95 KB

Response Body (JSON):

```
{
  "metadata": {
    "metadato": {
      "codigo": "28093c39-decf-462d-adfb-251e8cef5650",
      "idioma": "spa",
      "codificacion": "utf8",
      "contacto": {
        "rol": "author",
        "organizacion": "Instituto Geográfico Agustín Codazzi- Subdirección de Agrología - Grupo Interno de Trabajo Geomática -",
        "cargo": "Subdirector de Agrología",
        "telefono": {
          "numero": "+57 1 3694100 Ext. 91007",
          "tipo": null
        }
      },
      "direccion": "Carrera 30 # 48 - 51 - Sede Central",
      "ciudad": "Bogotá D.C",
      "pais": "República de Colombia",
      "email": "german.alvarez@idac.gov.co"
    }
  }
}
```

METADATOS - VALIDACIÓN

Geoservicio

POST ▼ https://rzzcm7sqlgbtuw6b6azoyhp7me0xbwvd.la

Send ▼

200 OK

3.43 s

3.5 KB

Params

Body ●

Auth ●

Headers 5

Scripts

Docs

Preview

Headers 6

Cookies

URL PREVIEW

https://rzzcm7sqlgbtuw6b6azoyhp7me0xbwvd.lambda-url.us-east-1.on.aws/?url=https%3A%2F%2Fgeoportal.dane.gov.co%2Fmparcgis%2Frest%2Fservices%2FDivipola%2FServ_PuntosCentrosPoblados2013I%2FMapServer%2FWFSServer&type=WFS

Preview ▼

1 ▼ {

2 ▼ "metadata": {

3 ▼ "metadato": {

4 ▼ "codigo": "b66cd407-b

METADATOS - VALIDACIÓN

Vector, Raster

```
{
  "metadata": {
    "metadato": { ... },
    "identificacion": { ... },
    "vector_info": { ... },
  },
  "validation": {
    "metrics": { ... },
    "fields": { ... },
    "validation_id": "20241124_232605"
  },
  "urls": { ... }
}
```

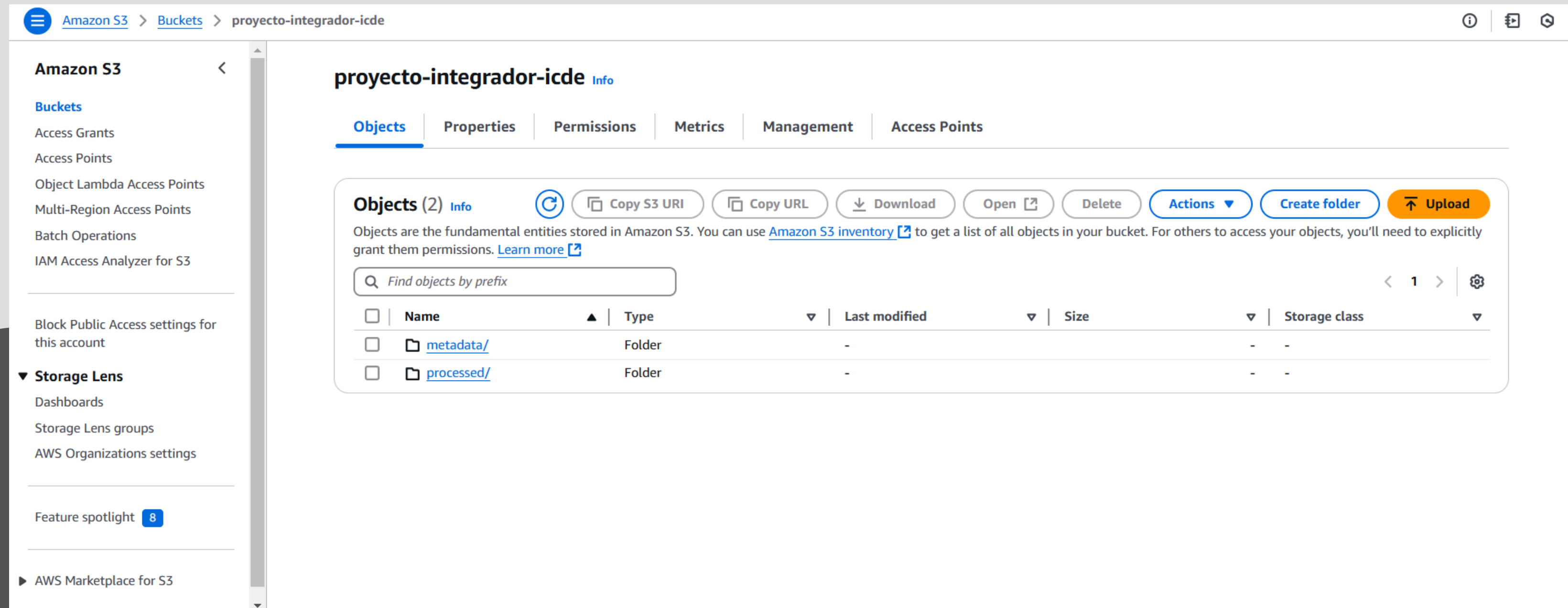
METADATOS - VALIDACIÓN

Geoservicio

```
1 ▾ {  
2 ▾   "metadata": {  
3 ▸     "metadato": { ← 8 → },  
35 ▸     "identificacion": { ← 9 → }  
65   },  
66 ▸   "validation": { ← 6 → },  
108 ▸   "urls": { ← 3 → },  
113 ▸   "live_check": { ← 2 → }  
121 }
```

METADATOS - VALIDACIÓN

Geoservicio, Vector, Raster












The screenshot displays the Amazon S3 console interface for a bucket named 'projecto-integrador-icde'. The left sidebar shows the navigation menu with 'Amazon S3' selected. The main content area has tabs for 'Objects', 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'. The 'Objects' tab is active, showing a list of two objects: 'metadata/' and 'processed/'. Above the list, there are buttons for 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', 'Actions', 'Create folder', and 'Upload'. A search bar is present above the table. The table has columns for 'Name', 'Type', 'Last modified', 'Size', and 'Storage class'.

Amazon S3 > Buckets > proyecto-integrador-icde



projecto-integrador-icde Info

Objects Properties Permissions Metrics Management Access Points

Objects (2) Info   Copy S3 URI  Copy URL  Download  Open  Delete  Actions  Create folder  Upload

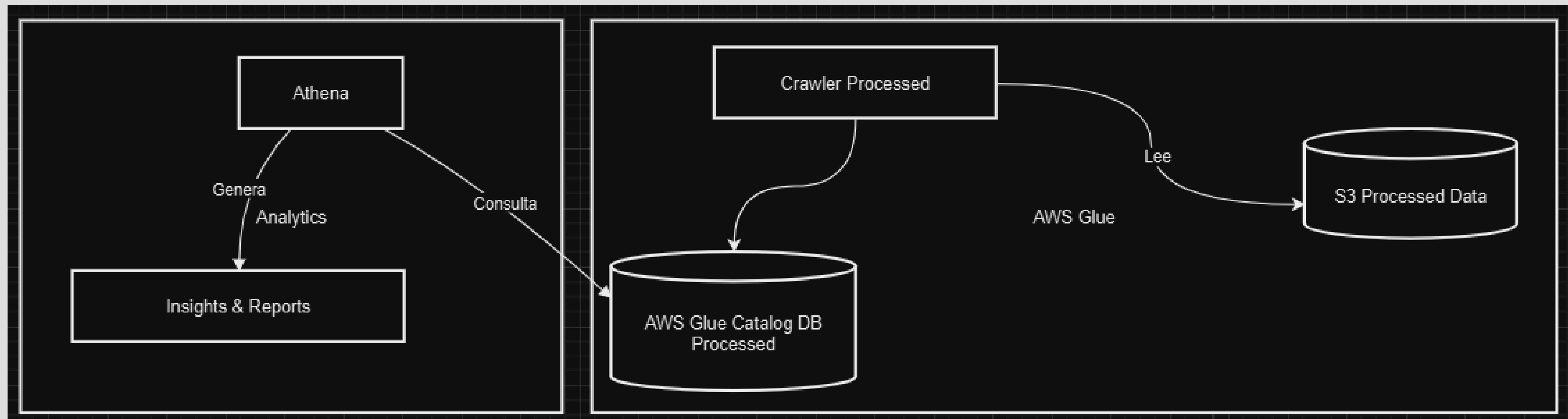
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 metadata/	Folder	-	-	-
<input type="checkbox"/>	 processed/	Folder	-	-	-

CICLO DE ANALÍTICA

Metadatos para
Geoservicio, Vector, Raster



METADATOS - OLAP

AWS Glue

Crawlers

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

▼ Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

► Data Integration and ETL

► Legacy pages

What's New

Documentation

AWS Marketplace

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (2) Info

Last updated (UTC)
November 24, 2024 at 23:32:33

Action

Run

Create crawler

View and manage all available crawlers.

Filter crawlers

< 1 >

<input type="checkbox"/>	Name	State	Schedule	Last run	Last run timestamp	Log	Table changes from l...
<input type="checkbox"/>	validation field icde	✓ Ready		✓ Succeeded	November 21, 2024 a...	View log	1 created
<input type="checkbox"/>	validation metrics icde	✓ Ready		✓ Succeeded	November 21, 2024 a...	View log	1 updated

METADATOS - OLAP

- Top Campos Problemáticos por Tipo
- Análisis de Secciones con Más Problemas
- Análisis Mensual de Completitud
- Distribución de Calidad por Tipo
- Estadísticas por Tipo de Metadato

Query 1

```
1 SELECT
2   DATE_TRUNC('month', parse_datetime(SUBSTRING(timestamp, 1, 19), 'yyyy-MM-dd''T''HH:mm:ss')) as mes,
3   metadata_type,
4   AVG(CAST(completeness_percentage AS double)) as promedio_completitud,
5   COUNT(*) as total_registros
6 FROM "icde"."metrics"
7 GROUP BY
8   DATE_TRUNC('month', parse_datetime(SUBSTRING(timestamp, 1, 19), 'yyyy-MM-dd''T''HH:mm:ss')),
```

SQLLn 10, Col 19

Ejecutar de nuevo

Explicar

Cancelar

Borrar

Crear

Volver a utilizar los resultados de la consulta hasta hace 60 minutos

Resultados de la consulta

Estado de la consulta

Completado

Tiempo en cola: 104 msTiempo de ejecución: 1.07 secDatos analizados: 4.75 KB

Copiar

Descargar resultados

Resultados (2)

Filas de búsqueda

#	mes	metadata_type	promedio_completitud	total_registros
1	2024-11-01 00:00:00.000 UTC	grid	77.78	4
2	2024-11-01 00:00:00.000 UTC	vector	82.96666666666663	24

METADATOS - OLAP

Query 1

1 SELECT

2 SPLIT_PART(f.field_name, '.', 1) as seccion_principal,

3 m.metadata_type,

4 COUNT(*) as total_campos,

5 SUM(CASE WHEN f.field_status = 'missing' THEN 1 ELSE 0 END) as campos_faltantes,

6 ROUND(SUM(CASE WHEN f.field_status = 'missing' THEN 1 ELSE 0 END) * 100.0 / COUNT(*), 2) as porcentaje_faltante

7 FROM "icde"."fields" f

8 JOIN "icde"."metrics" m

SQL Ln 14, Col 35

Ejecutar de nuevo

Explicar

Cancelar

Borrar

Crear

☐ Volver a utilizar los resultados de la consulta hasta hace 60 minutos

Resultados de la consulta

Estado de la consulta

Completado

Tiempo en cola: 111 ms Tiempo de ejecución: 1 sec Datos analizados: 93.22 KB

Resultados (6)

Copiar Descargar resultados

Filas de búsqueda

< 1 >

#	seccion_principal	metadata_type	total_campos	campos_faltantes	porcentaje_faltante
1	metadato	grid	76	24	31.58
2	metadato	vector	456	92	20.18
3	identificacion	grid	88	16	18.18
4	identificacion	vector	348	48	13.79
5	vector_info	vector	48	4	8.33
6	raster_info	grid	16	0	0.0

METADATOS - OLAP

Query 1

1 SELECT

2 DATE_TRUNC('month', parse_datetime(SUBSTRING(timestamp, 1, 19), 'yyyy-MM-dd' 'T' 'HH:mm:ss')) as mes,

3 metadata_type,

4 AVG(CAST(completeness_percentage AS double)) as promedio_completitud,

5 COUNT(*) as total_registros

6 FROM "icde"."metrics"

7 GROUP BY

8 DATE_TRUNC('month', parse_datetime(SUBSTRING(timestamp, 1, 19), 'yyyy-MM-dd' 'T' 'HH:mm:ss')),

SQL Ln 10, Col 19

Ejecutar de nuevo

Explicar

Cancelar

Borrar

Crear

☐ Volver a utilizar los resultados de la consulta hasta hace 60 minutos

Resultados de la consulta

Estado de la consulta

Completado

Tiempo en cola: 104 ms Tiempo de ejecución: 1.07 sec Datos analizados: 4.75 KB

Resultados (2)

Copiar

Descargar resultados

Filas de búsqueda

< 1 >

#	mes	metadata_type	promedio_completitud	total_registros
1	2024-11-01 00:00:00.000 UTC	grid	77.78	4
2	2024-11-01 00:00:00.000 UTC	vector	82.96666666666663	24

DATOS - TIPOS DE DATOS

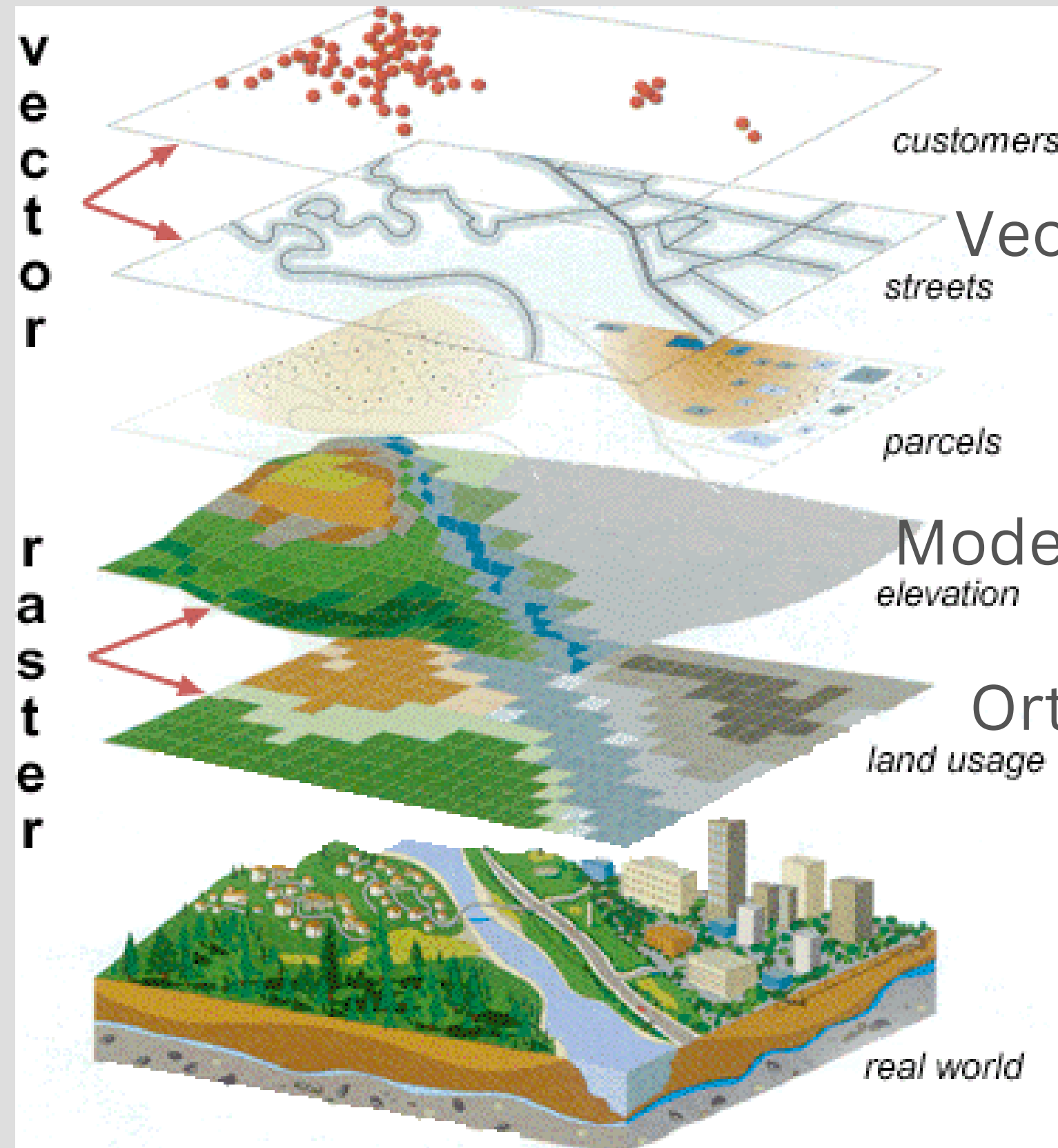
Vectorial

- Base de datos(GDB)
- Poligonos
- Lineas
- Punto
- Formatos
 - gdb
 - gpkg
 - shp
 - kml
 - dnf

Raster

- Modelos Digitales
- Ortoimagenes
- Formatos
 - TIFF
 - GEOTIFF

DATOS - COMO FUNCIONAN

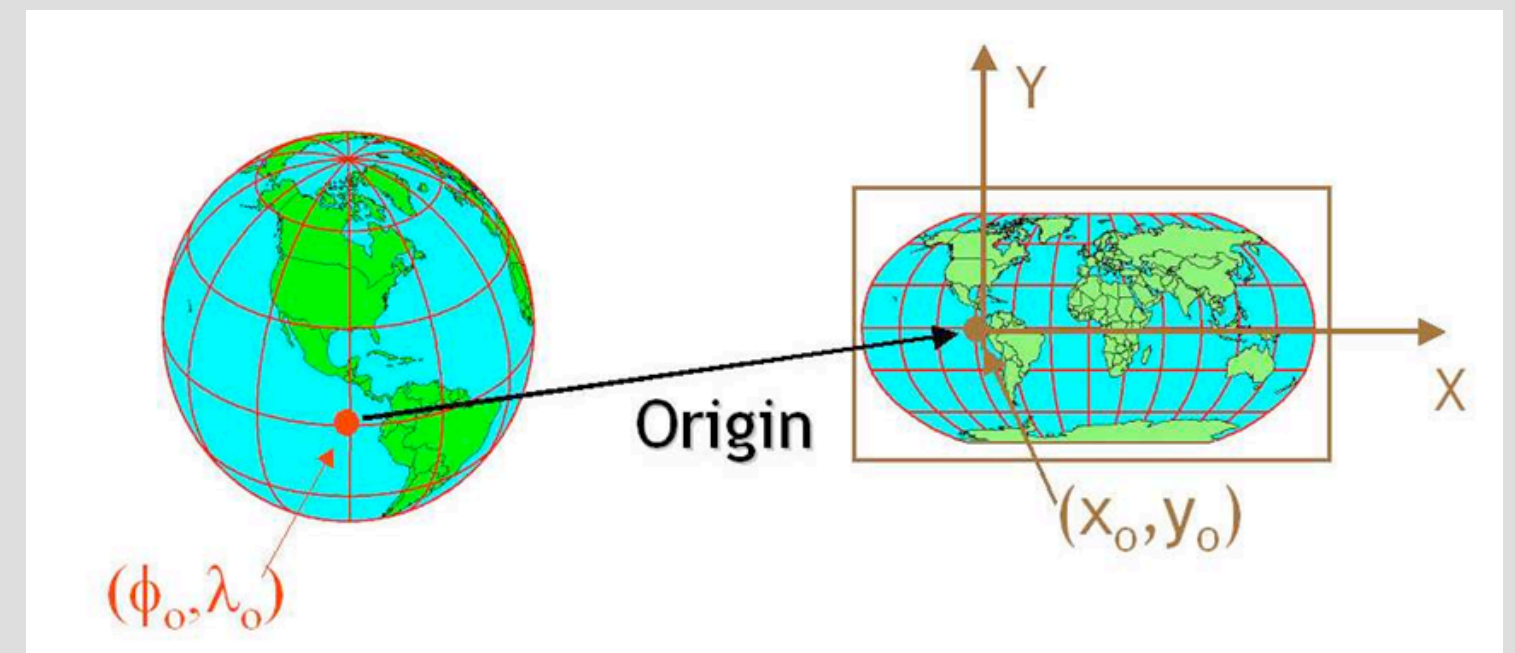
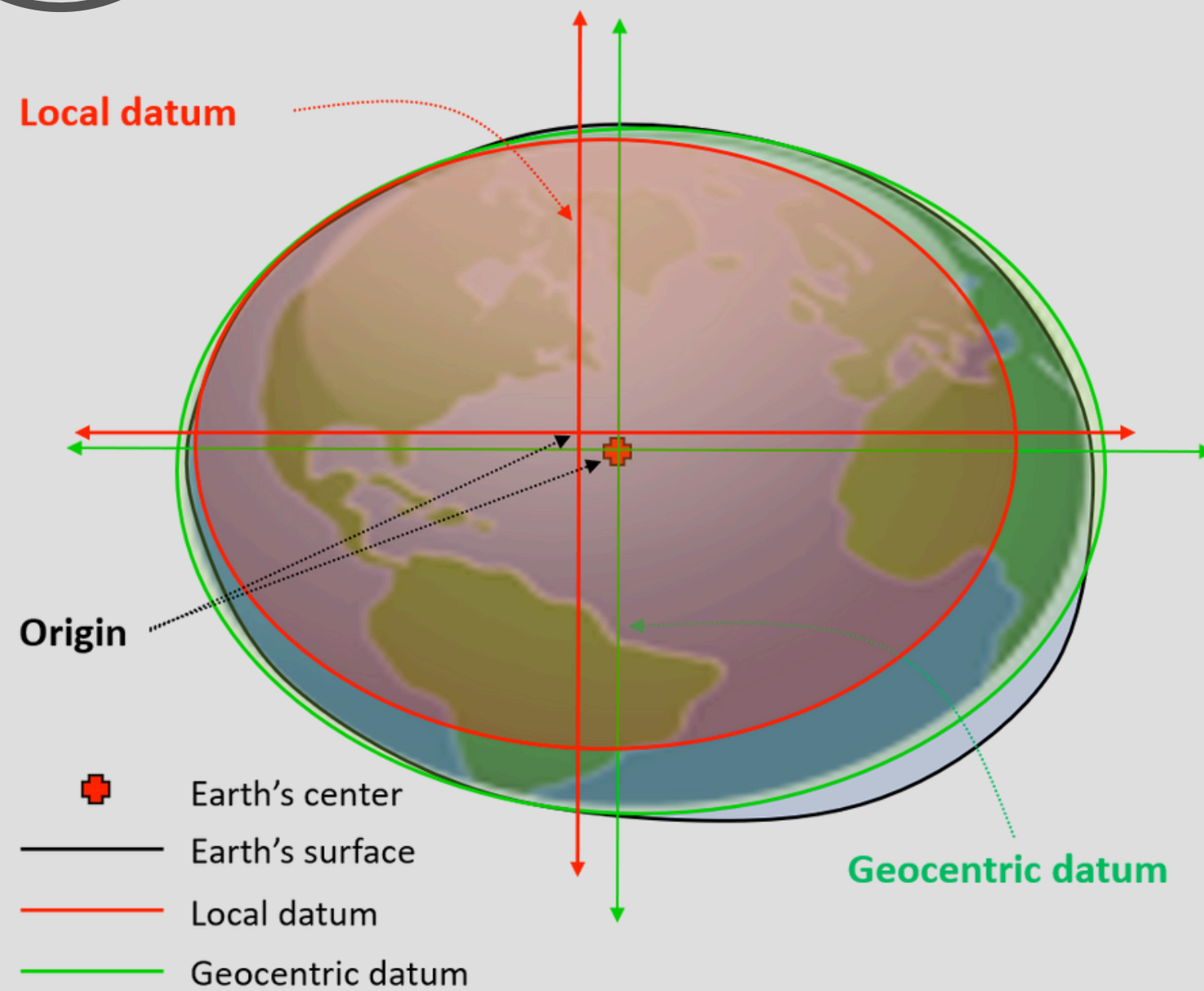


Vectores

Modelos digitales

Ortoimágenes

DATOS - COMO FUNCIONAN



DATOS - VALIDACIÓN

Conceptual

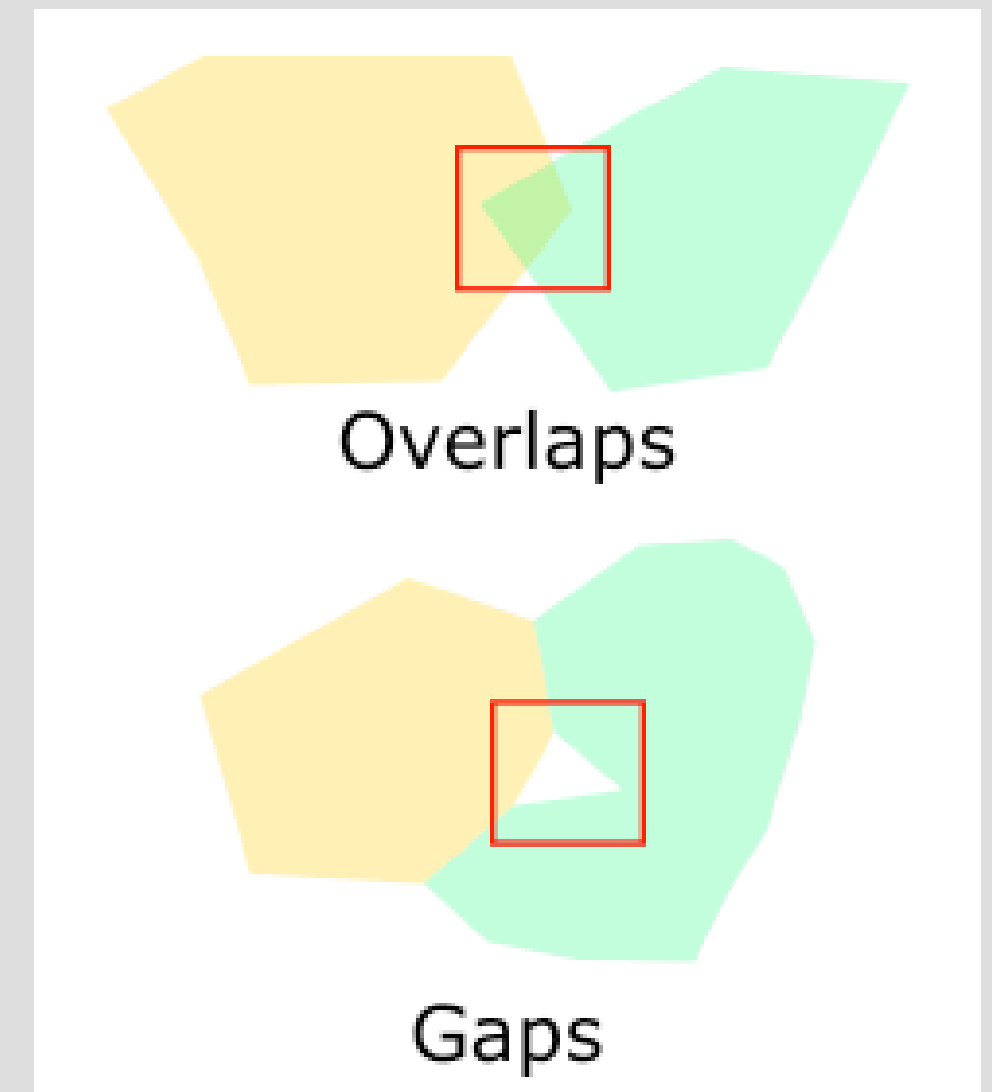
- Dentro de Colombia
- Unico punto de origen
 -

Formato

- Formato de archivo
- Ortoimagen
 - Colores de bandas
 - Radiometria

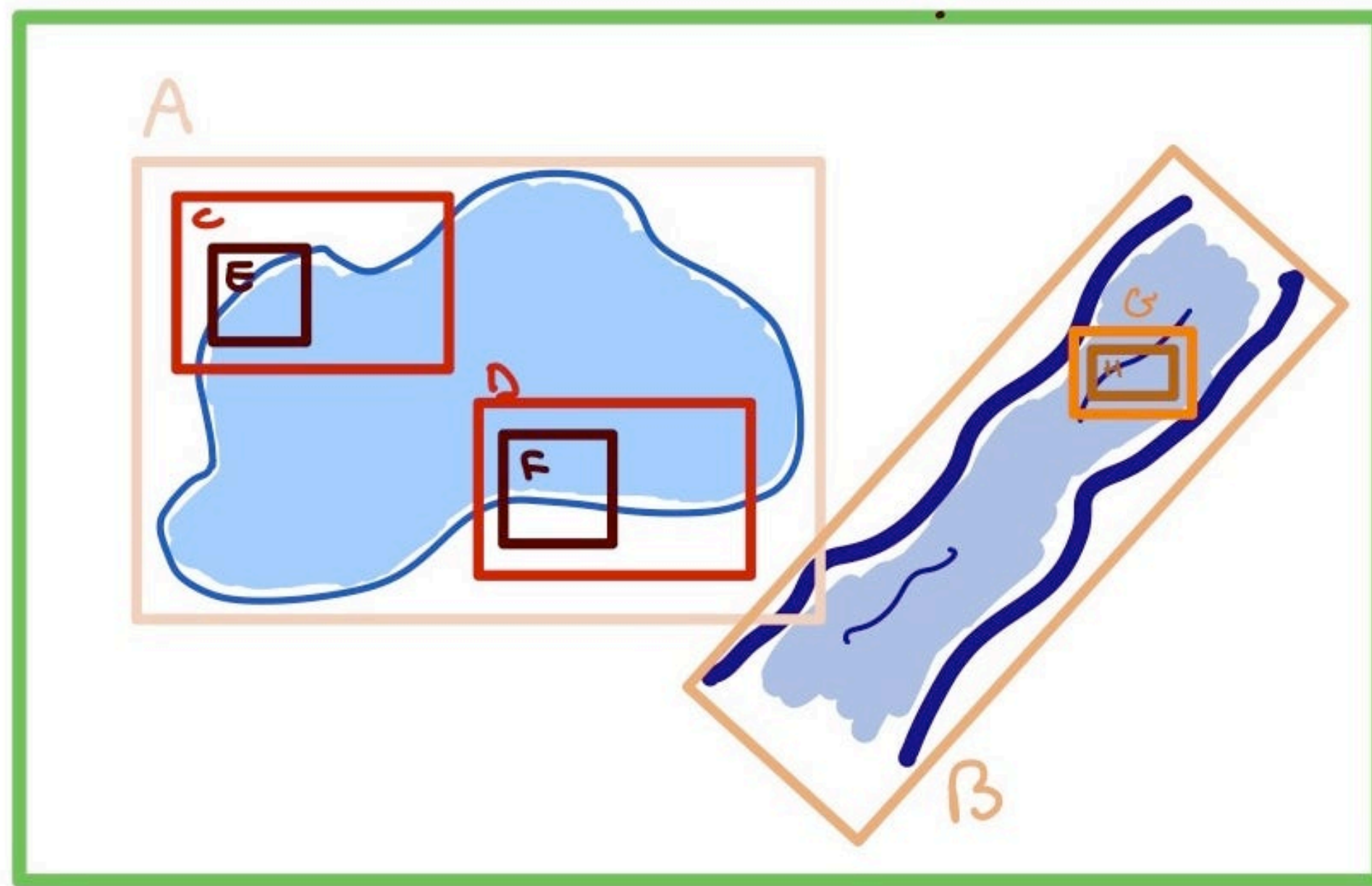
Topologica

- Gaps
- Superposición



DATOS - SUPERPOSICIÓN

STRTREE



DATOS - PROGRAMAS



GeoPandas



docker

DATOS - OLAP

Vector

	valores_nulos	dentro_de_origen	consistencia_de_orig	hueco_en_capa	superposicion
count	34.0	34.0	34.0	34.0	34.0
mean	23.735	0.0	0.441	3.0	0.971
std	7.004	0.0	0.504	1.792	0.171
min	1.0	0.0	0.0	0.0	0.0
25%	20.0	0.0	0.0	2.0	1.0
50%	24.0	0.0	0.0	3.0	1.0
75%	27.0	0.0	1.0	4.0	1.0
max	45.0	0.0	1.0	8.0	1.0
percent_error	1.0	0.0	0.441	0.853	0.971

DATOS - OLAP

Modelos Digitales

	dentro_de_origen	consistencia_de_orig
count	30.0	30.0
mean	0.0	0.833
std	0.0	0.379
min	0.0	0.0
25%	0.0	1.0
50%	0.0	1.0
75%	0.0	1.0
max	0.0	1.0
percent_error	0.0	0.833

DATOS - OLAP

Ortoimagenes

Unnamed: 0	dentro_de_origen	consistencia_de_orig	bandas	radiometria
count	11.0	11.0	11.0	11.0
mean	0.0	0.545	0.182	0.0
std	0.0	0.522	0.405	0.0
min	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0
50%	0.0	1.0	0.0	0.0
75%	0.0	1.0	0.0	0.0
max	0.0	1.0	1.0	0.0
percent_error	0.0	0.545	0.182	0.0

ETAPAS DEL CICLO DE VIDA

1. Ingesta:

- Recepción de datos crudos.
- Almacenamiento en un Data Lake (Amazon S3).

2. Preparación:

- Limpieza y transformación mediante AWS Glue.
- Organización en tablas estructuradas para consultas SQL.

3. Procesamiento y Validación:

- Verificación de criterios:
 - Consistencia lógica.
 - Conformidad de metadatos.
 - Resolución espacial.
- Automatización con AWS Lambda y EC2 Docker.

RESULTADOS Y ALMACENAMIENTO

Resultados:

- Validación en tiempo real mediante APIs REST.
- Generación de reportes descargables.

Almacenamiento y Consulta:

- Tablas estructuradas accesibles con Amazon Athena.
- Consultas flexibles para análisis posteriores.

BENEFICIOS DEL CICLO DE VIDA AUTOMATIZADO

Elimina errores manuales.

Escalabilidad y eficiencia.

Datos listos para análisis en entornos críticos
(urbanismo, emergencias, etc.).

GRACIAS