

Modernización tecnológica para facilitar la analítica de información catastral

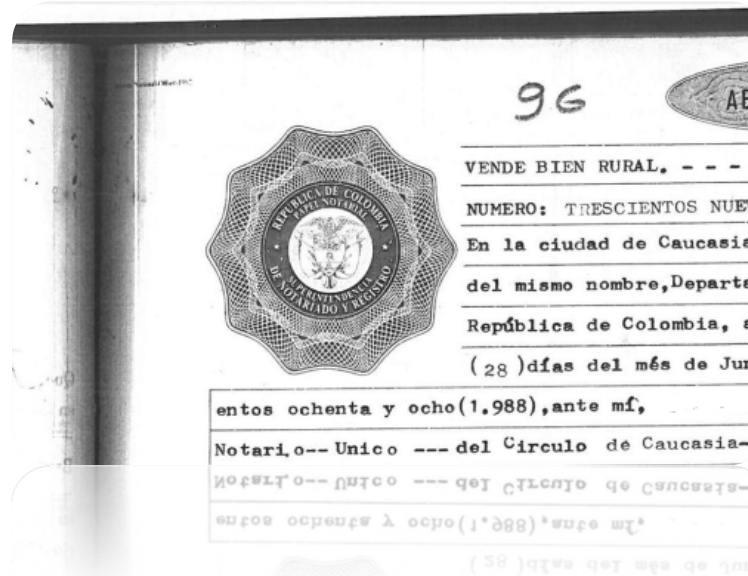
*Revolución de datos en el catastro con enfoque
multipropósito - Simposio CPCI*

Juan Carlos Méndez

*Arquitecto de Sistemas de Información Geográfica
y Nube Pública - IGAC*
juan.mendez@igac.gov.co

El ***catastro multipropósito*** en Colombia integra la ***ciencia de datos*** como parte de la política pública para mejorar la gestión de la tierra. El decreto 148 de 2020 contempla dentro de los “Procedimientos de Enfoque Multipropósito” el ***uso de métodos directos e indirectos*** para la recolección de información catastral, apoyados en ***análisis de Big Data***, econometría y ***tecnología***.

Métodos indirectos: Son aquellos métodos de identificación física, jurídica y económica de los bienes inmuebles a través del uso de imágenes de sensores remotos, integración de registros administrativos, modelos estadísticos y econométricos, análisis de Big Data y demás fuentes secundarias como los observatorios inmobiliarios, para su posterior incorporación en la base catastral.



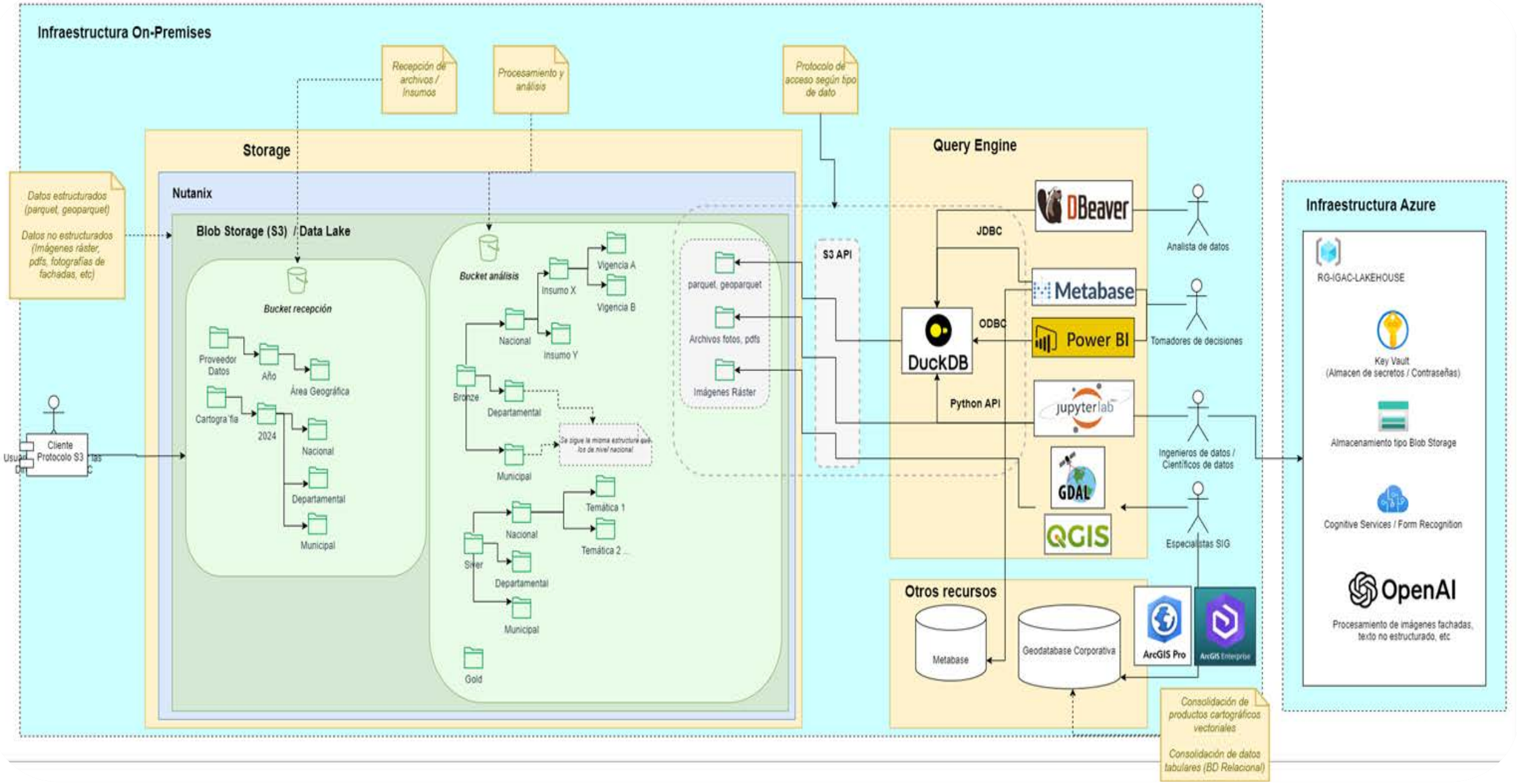
Tecnología

Oportunidad
Análisis **Datos**
Volúmen Eficiencia
Diversidad



Estándares
Arquitectura híbrida
Roles Integración
Lago de datos
Nube pública

Infraestructura On-Premises



- Geoparquet v1.1.0: Extensión del formato de archivo **columnar Parquet**, optimizado para almacenar datos geoespaciales de manera eficiente y escalable (**OGC incubating**)
- Este tipo de formato facilitan la ejecución de cargas de trabajo de tipo analítico (**OLAP – Online Analytical Processing**) para grandes volúmenes de datos

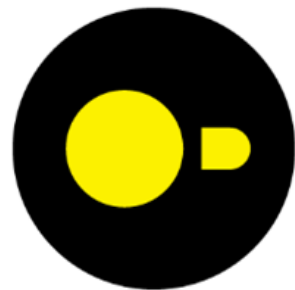
	session_id	timestamp	source_ip
Row 1	1331246660	3/8/2012 2:44PM	99.155.155.225
Row 2	1331246351	3/8/2012 2:38PM	65.87.165.114
Row 3	1331244570	3/8/2012 2:09PM	71.10.106.181
Row 4	1331261196	3/8/2012 6:46PM	76.102.156.138

	1331246660
Row 1	3/8/2012 2:44PM
	99.155.155.225
	1331246351
Row 2	3/8/2012 2:38PM
	65.87.165.114
	1331244570
Row 3	3/8/2012 2:09PM
	71.10.106.181
	1331261196
Row 4	3/8/2012 6:46PM
	76.102.156.138

	1331246660
session_id	1331246351
	1331244570
	1331261196
timestamp	3/8/2012 2:44PM
	3/8/2012 2:38PM
	3/8/2012 2:09PM
	3/8/2012 6:46PM
source_ip	99.155.155.225
	65.87.165.114
	71.10.106.181
	76.102.156.138

SELECT * FROM clickstream
WHERE session_id = 1331246





DuckDB

DuckDB is a fast in-process analytical database

DuckDB supports a feature-rich SQL dialect
complemented with deep integrations into client APIs.

Installation ↓

Documentation

SQL

Python

R

Java

Node.js

```
1  -- Directly query Parquet file in S3
2  SELECT station_name, count(*) AS num_services
3  FROM 's3://duckdb-blobs/train_services.parquet'
4  GROUP BY ALL
5  ORDER BY num_services DESC
6  LIMIT 10;
```

Read Parquet files from S3



Live demo →

1,761,092,498

Total de registros cargados en el Data Lake ⓘ

Total de registros por tipo

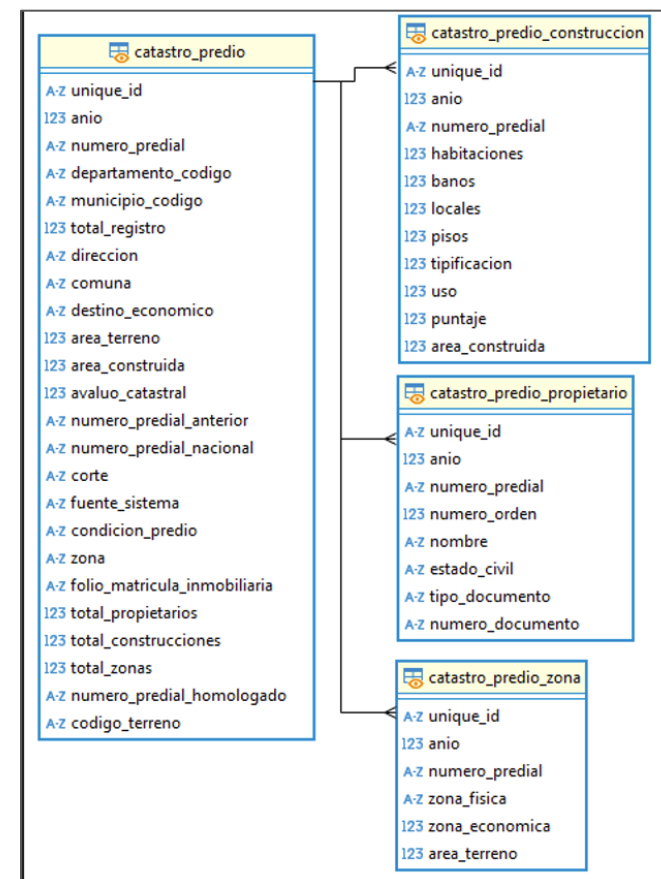
Tipo Dato	Total
ALFANUMERICO	1,477,691,799
GEOGRAFICO	283,400,699
Grand totals	1,761,092,498

PREDIO	227,434,773
PREDIO_CONSTRUCCION	184,007,977
PREDIO_PROPIETARIO	300,660,902
PREDIO_R1	300,660,912
PREDIO_R2	235,601,980
PREDIO_ZONA	229,325,255

GEO_R_CONSTRUCCION	3,979,067
GEO_R_CONSTRUCCION_INFORMAL	96,645
GEO_R_NOMENCLATURA_DOMICILIARIA	1,328,342
GEO_R_NOMENCLATURA_VIAL	256,261
GEO_R_SECTOR	26,115
GEO_R_TERRENO	41,921,377
GEO_R_TERRENO_INFORMAL	76,120
GEO_R_UNIDAD	3,506,026
GEO_R_UNIDAD_INFORMAL	105,126
GEO_R_VEREDA	251,008
GEO_R_ZONA_HOMOGENEA_FISICA	472,065
GEO_R_ZONA_HOMOGENEA_GEOECONOMICA	975,923
GEO_U_CONSTRUCCION	77,891,203
GEO_U_CONSTRUCCION_INFORMAL	47,717
GEO_U_MANZANA	3,343,982
GEO_U_NOMENCLATURA_DOMICILIARIA	48,508,424
GEO_U_NOMENCLATURA_VIAL	3,032,417
GEO_U_PERIMETRO	35,106
GEO_U_SECTOR	29,009
GEO_U_TERRENO	63,068,417
GEO_U_TERRENO_INFORMAL	35,397

Total de registros por entidad y año

Año	PREDIO	PREDIO_CONSTRUCCION	PREDIO_PROPIETARIO	PREDIO_R1	PREDIO_R2	PREDIO_ZONA	Total
2000	7.463.191	5.193.005	9.785.544	9.785.544	7.563.558	7.248.311	47.039.153
2001	7.614.876	5.340.666	9.992.535	9.992.535	7.718.041	7.395.299	48.053.952
2002	7.718.287	5.397.448	10.149.966	10.149.966	7.821.910	7.508.242	48.745.819
2003	7.901.083	5.573.716	10.402.005	10.402.005	8.008.562	7.689.197	49.976.568
2004	8.117.550	5.786.165	10.700.105	10.700.105	8.229.456	7.915.419	51.448.800
2005	8.160.246	5.869.546	10.752.928	10.752.928	8.274.108	7.972.638	51.782.394
2006	8.414.276	6.243.731	11.089.448	11.089.448	8.540.197	8.225.176	53.602.276
2007	8.610.580	6.512.528	11.344.233	11.344.233	8.747.638	8.411.613	54.970.825
2008	8.806.558	6.798.335	11.619.512	11.619.514	8.955.890	8.605.705	56.405.514
2009	8.962.155	7.127.809	11.868.899	11.868.899	9.142.693	8.829.495	57.799.950
2010	9.260.900	7.481.315	12.259.191	12.259.191	9.460.778	9.147.215	59.868.590
2011	9.492.470	7.838.884	12.581.472	12.581.472	11.175.926	9.401.116	63.071.340
2012	9.707.615	8.190.310	12.868.152	12.868.152	9.963.840	9.667.437	63.265.506
2013	9.913.736	8.474.796	13.147.728	13.147.728	10.198.916	9.945.513	64.828.417
2014	10.178.126	8.848.107	13.480.432	13.480.432	10.524.082	10.341.887	66.853.066
2015	10.368.430	9.118.709	13.728.558	13.728.559	10.793.024	10.581.925	68.319.205
2016	10.580.844	9.280.691	14.006.458	14.006.458	10.982.503	10.880.402	69.737.356
2017	10.851.447	9.506.125	14.373.901	14.373.904	11.272.672	11.192.477	71.570.526
2018	11.055.140	9.698.625	14.688.806	14.688.809	11.499.616	11.451.424	73.082.420
2019	11.274.063	9.813.042	14.957.696	14.957.697	11.746.306	11.247.228	73.996.032
2020	11.501.429	10.115.174	15.286.464	15.286.464	11.985.076	12.018.292	76.192.899
2021	10.000.196	8.666.194	13.165.697	13.165.697	10.452.644	10.529.048	65.979.476
2022	7.793.760	6.327.619	10.259.662	10.259.662	8.140.234	8.258.749	51.039.686
2023	6.923.016	5.475.594	9.162.164	9.162.164	7.269.212	7.532.727	45.524.877
2024	6.764.799	5.329.843	8.989.346	8.989.346	7.135.098	7.328.720	44.537.152
Total	227.434.773	184.007.977	300.660.902	300.660.912	235.601.980	229.325.255	1.477.691.799



659 SELECT periodo_anio, count(*)

660 FROM read_parquet('s3://objectspruebas/datos/catastro/geo/**/u_terreno*.parquet', hive_partitioning = true

661 GROUP BY periodo_anio

662 ORDER BY periodo_anio

Results 1

SELECT periodo_anio, count(*) Enter a SQL expression to filter results (use Ctrl+Space)

	123 periodo_anio	123 count_star()
1	2,010	4,330,733
2	2,011	4,292,519
3	2,012	3,214,059
4	2,013	4,548,489
5	2,014	4,877,417
6	2,015	4,997,629
7	2,016	4,569,599
8	2,017	4,890,293
9	2,018	4,972,859
10	2,019	4,967,699
11	2,020	4,647,803
12	2,022	3,009,060

SQL Commit Rollback Auto testblob.db testblob

Query Manager

Type query part to search in query history

Time	Type	Text	Duration (ms)	Rows
Aug-15 17:...	SQL / User	SELECT periodo_anio, count(*) FROM read_...	5,069	12

Conteo de registros en la entidad Predio derivada de R1 / R2



Código departamento

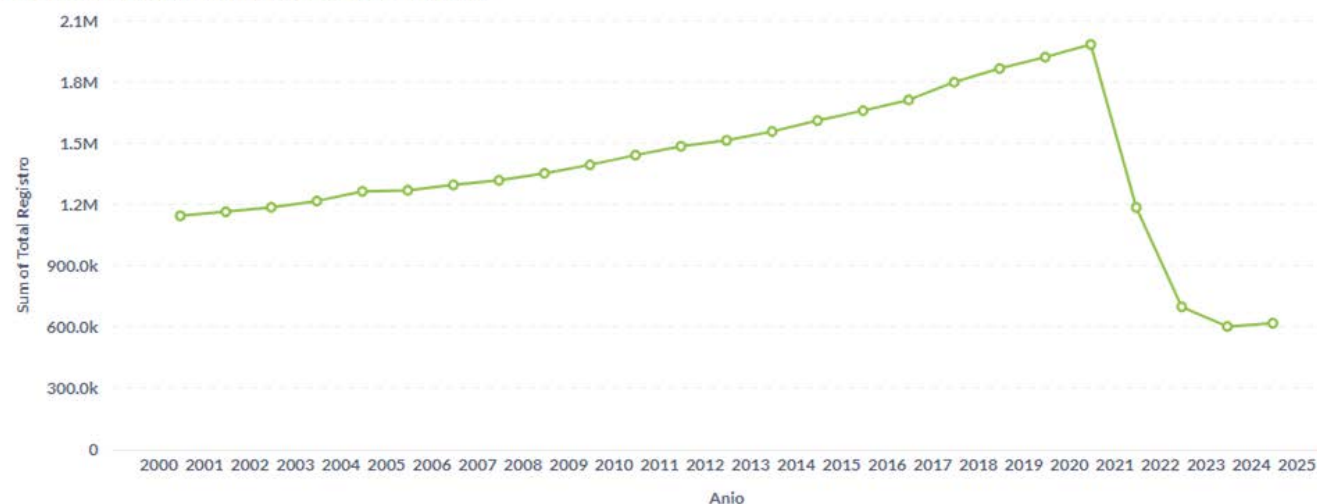
25



Código municipio



Catastro Predio, Sum of Total Registro, Grouped by Anio



See these Catastro Predios

Zoom in

Break out by...

Automatic insights...

Filter by this value



2019 2020 2021 2022 2023 2024 2025

Tipo de dato
GEOGRAFICO

283.4M

Total de registros cargados en
el Data Lake

Total de datos cargados por conjunto de datos y año

	IGAC					
	BRONZE					
	CATASTRO					
	GEOGRAFICO					
Anio	GEO_R_CONSTRUCC...	GEO_R_CON...	GEO_R_NOM...	GEO_R_NOM...	GEO_R_SECT...	GEO_R_TERR...
2010	34,526					
2011	62,540					
2012	26,549		1,905	1,138	201	
2013	194,102		69,337	36,964	1,791	
2014	278,740		77,310	35,885	1,877	
2015	290,700		73,962	9,483	2,010	
2016	258,335		66,249	19,507	1,930	
2017	317,958		98,500	19,691	2,025	
2018	371,495		108,819	20,410	2,647	
2019	402,150		119,897	20,520	2,645	
2020	426,011		124,449	20,437	2,489	
2021	355,602		104,547	19,521	2,247	
2022	269,899	9,719	123,969	17,844	2,139	
2023	316,505	17,536	165,041	17,397	1,861	
2024	373,955	69,390	194,357	17,464	2,253	
Grand t...	3,979,067	96,645	1,328,342	256,261	26,115	

```
664 INSTALL spatial;
665 LOAD spatial;
666 |
667 SELECT codigo_municipio, codigo, periodo_anio,
668 ST_Transform(st_centroid(ST_GeomFromWKB(wkt)), 'EPSG:9377', 'EPSG:4326') AS centroid
669 FROM read_parquet('s3://objectspruebas/datos/catastro/geo/**/u_terreno*.parquet', hive_partitioning = true)
670 WHERE periodo_anio = 2016 AND codigo_departamento = '25'
671 LIMIT 100;
```

results 1

ELECT codigo_municipio, codigo, periodo_anio, ST_Transform | Enter a SQL expression to filter results (use Ctrl+Space)

ABC codigo_municipio	ABC CODIGO	123 periodo_anio	centroid
25001	250010100000000200009000000000	2,016	POINT (26.291240381491278 -101.89124449108513)
25001	250010100000000200015000000000	2,016	POINT (26.2912963927846 -101.89182958821911)
25001	250010100000000200014000000000	2,016	POINT (26.291268380147812 -101.89173347996159)
25001	250010100000000200013000000000	2,016	POINT (26.291260860002176 -101.89158484037253)
25001	250010100000000200012000000000	2,016	POINT (26.29122726730333 -101.89153188861349)
25001	250010100000000200016000000000	2,016	POINT (26.291329919625852 -101.89166844454175)
25001	250010100000000200017000000000	2,016	POINT (26.291400734782183 -101.89160677975175)
25001	250010100000000200018000000000	2,016	POINT (26.291449759725776 -101.8916440644322)
25001	250010100000000200036000000000	2,016	POINT (26.29146141895394 -101.89157758746404)
25001	250010100000000200019000000000	2,016	POINT (26.291510797566662 -101.89162890127393)
25001	250010100000000200020000000000	2,016	POINT (26.29150707067206 -101.8915310004374)

Refresh Save Cancel | Export data 200 100 | 100 row(s) fetched

Avalúos Comerciales OIM



Departamento ▼ Municipio ▼ Fecha Avalúo ▼

698,080

Total de registros ⓘ

33

Total Departamentos ⓘ

1,046

Total Municipios ⓘ

645,419

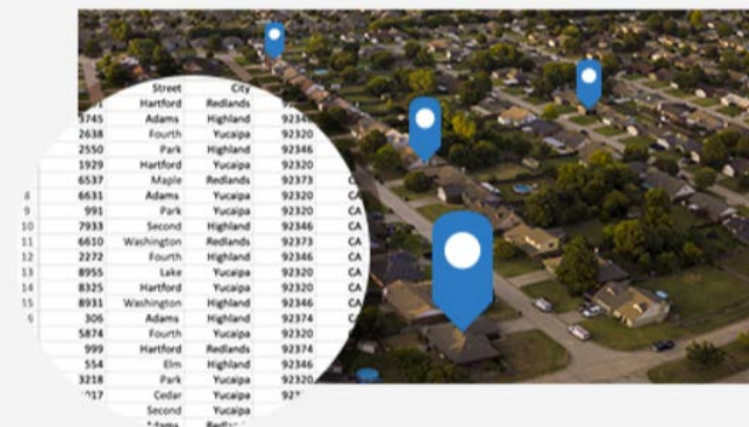
Total registros georreferenciados ⓘ

Oim Avaluos Comerciales, Count, Grouped by Departamento, Municipio y Año del Avalúo

Departamento	Municipio	2020	2021	2022	2023	2024	Row totals
Totals for AMAZONAS		10	19	12			42
Totals for ANTIOQUIA		4,416	12,100	17,426	3,647	3	37,752
Totals for ARAUCA		590	1,064	1,584	359		3,609
Totals for ARCHIPIÉLAGO DE SAN ANDRÉS, PR...		7	43	155	16		226
Totals for ATLÁNTICO		203	393	407	38		1,043
Totals for BOGOTÁ, D.C.		9,021	48,903	124,259	54,966		237,598
Totals for BOLÍVAR		434	1,493	3,506	705		6,141
Totals for BOYACÁ		2,075	5,146	5,958	1,043		14,228
Totals for CALDAS		1,162	3,303	6,523	1,492		12,519
Totals for CAQUETÁ		790	1,317	1,336	161		3,609
Totals for CASANARE		505	1,492	2,423	633		5,054
CAUCA	ALMAGUER		2	1			3
	ARGELIA	4	3	4			11
	BALBOA	7	11	9	1		28
	BOLÍVAR	2	5	12			19
	BUENOS AIRES	3	3				7
	CAJIBÍO	8	12	3	1		24

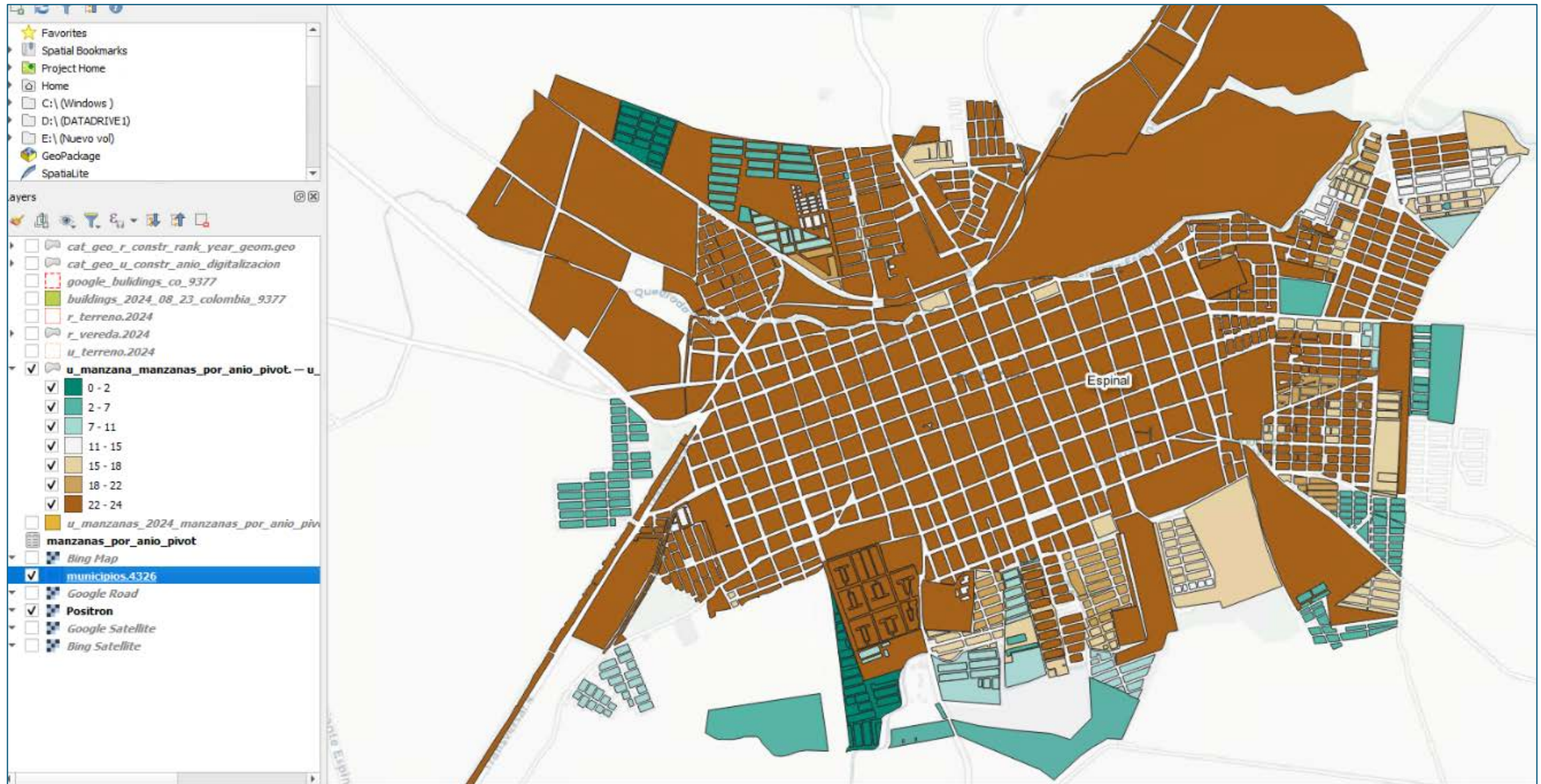
ArcGIS StreetMap Premium

Datos de calles de calidad tras su firewall para la visualización de mapas, geocodificación y generación de rutas



Geocodificación lista para usar

Analítica visual



Analítica visual



Integración de fuentes de datos externas



OVERTURE MAPS
FOUNDATION

Community Database License Agreement – Permissive v2
(CDLA)

Datasets

Microsoft Building Footprints

Google Research Open Buildings



OpenStreetMap



Integración de fuentes de datos externas

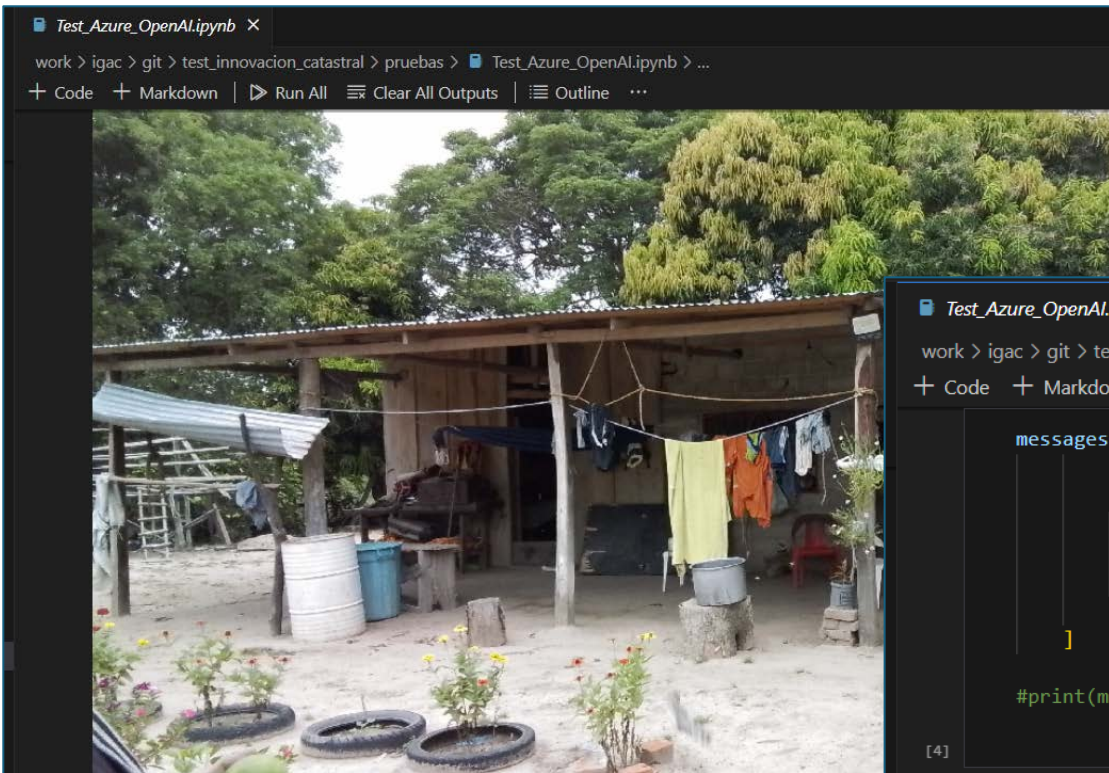


X

2,010
2,011
2,012
2,013
2,014
2,015
2,016
2,017
2,018
2,019
2,020
2,021
2,022
2,023
2,024



Inteligencia artificial - LLM



```
Test_Azure_OpenAI.ipynb X
work > igac > git > test_innovacion_catastral > pruebas > Test_Azure_OpenAI.ipynb > ...
+ Code + Markdown | ▶ Run All ⌵ Clear All Outputs | ⌵ Outline ...

Test_Azure_OpenAI.ipynb X
work > igac > git > test_innovacion_catastral > pruebas > Test_Azure_OpenAI.ipynb > ...
+ Code + Markdown | ▶ Run All ⌵ Clear All Outputs | ⌵ Outline ...

messages=[
    {"role": "system", "content": "You are a helpful assistant that responds in Markdown."},
    {"role": "user", "content": [
        {"type": "text", "text": "Como especialista en gestión catastral describa la imagen de manera detallada para identi"},
        {"type": "image_url", "image_url": { "url": f"data:image/png;base64,{base64_image}" , "detail": "high"}
    ]}
]

#print(messages)

[4]

ai_message = llm.invoke(messages)
print(ai_message.content)

[5]

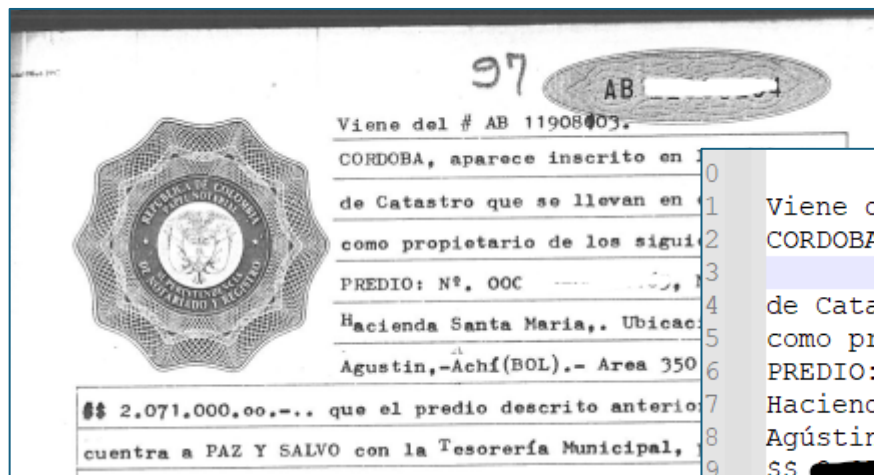
... La imagen muestra una construcción rural, posiblemente en una zona de clima cálido dado el tipo de vegetación y el diseño abierto de de

### Descripción Detallada de la Construcción:

1. **Materiales de Construcción**:
    - **Paredes**: Las paredes son de materiales mixtos, principalmente madera y bloques de concreto sin revocar, lo que sugiere una c
    - **Techo**: El techo está compuesto de láminas de zinc, algunas de las cuales están en condiciones deterioradas o desgastadas. Es

2. **Diseño y Layout**:
    - La casa tiene un diseño abierto con áreas que parecen estar destinadas para cocina y áreas de estar al aire libre. Esto es indio
    - Hay una estructura adicional con techo de lámina que podría servir como área de almacenamiento
```

Inteligencia artificial – OCR / LLM



Viene del f AB [REDACTED]
CORDOBA, aparece inscrito en los libros -
de Catastro que se llevan en este Mpio, -
como propietario de los siguientes predios
PREDIO: N°, 000, Nomenclatura
Hacienda Santa [REDACTED], Ubicación En San -
Agústín,-Achf(BOL) .us Area 350 Hst,- Avaluo
\$\$ [REDACTED].. que el predio descrito anteriormente
cuenta a PAZ Y SALVO con la Tesorería Municipal, por cond

6. ****Lindero****: La descripción completa de los linderos de la propiedad.

- ****Instrucciones****: Busca secciones que describan los límites con frases como "delimitado así", "colinda con". Devuelve la descripción
'null'.
- ****Ejemplos****:
 - "Delimitado el lote así: ORIENTE; deslinda con propiedad de Fredy Ruiz..." → (devuelve la descripción completa).
 - Sin mención → 'null'.

7. ****Escritura Pública****: Información de la escritura pública, incluyendo número, fecha y notaría.

- ****Instrucciones****: Busca frases como "ESCRITURA PÚBLICA NÚMERO", "Escritura No.", seguida de un número y fecha. Devuelve los detalles
'null'.

```
"content": {
  "departamento": "Cauca",
  "municipio": "Almaguer",
  "vereda": "Juan Ruiz",
  "cedula_catastral": "000 [REDACTED]",
  "matricula": "1 [REDACTED]",
  "lindero": "Por la cabecera linda con terrenos de la m",
  "escritura": "Escritura Pública No. [REDACTED] de octub",
  "area": "1,2500 mts²",
  "entidades": [
    "República de Colombia",
    "Notaría Única",
    "Superintendencia de Notariado y Registro",
    "Almaguer",
    "Cauca",
    "Bolívar",
    "María Ana",
    "Carmenza",
    "Gregorio",
    "Cira",
    "Tomás",
    "Cenen",
    "Adolfo",
    "Dora",
    "Registraduría Nacional del Estado Civil",
    "Escritura Pública No. [REDACTED]"
  ]
}
```

Gracias