El Poder de los datos en CEMEX y el arte de lo posible con el INEGI

snowflake®

Snowflake Team!

Paco Silva
Account Executive

Carlos Suárez
Solutions Engineer Enterprise

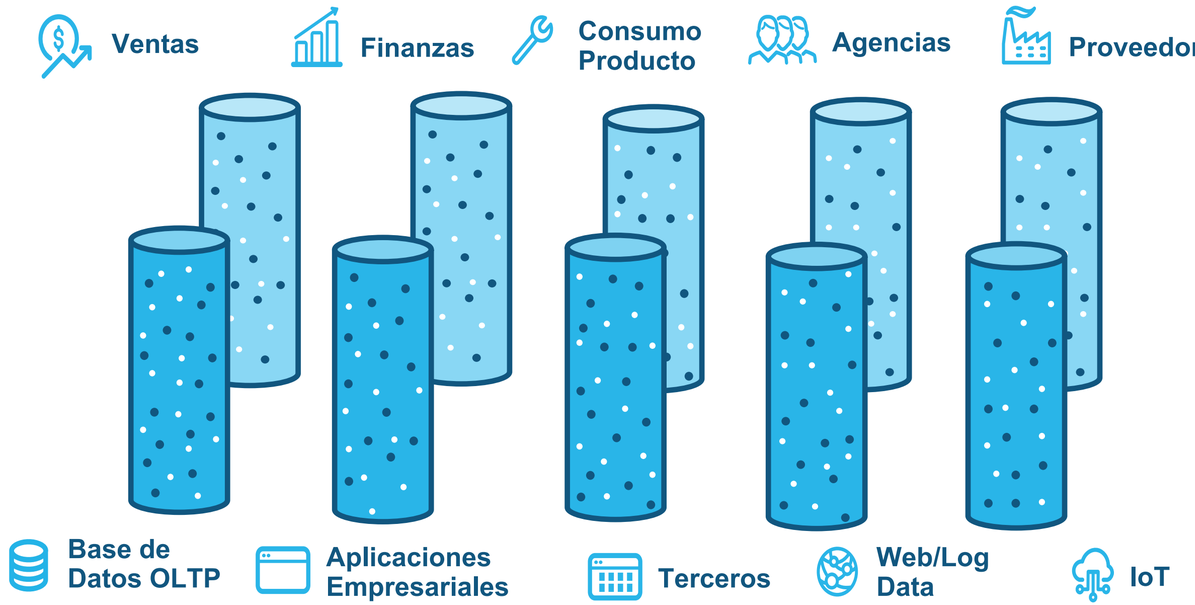
snowflake®

Realidad - Silos de Datos

El mundo real

FUENTES DATOS

- OLTP Bases de datos
- Aplicaciones Empresariales
- Terceros
- Web/Log Data
- IoT



CONSUMO DATOS

- Reportes de Operación
- Análisis
- Analítica en tiempo real

El Desafio

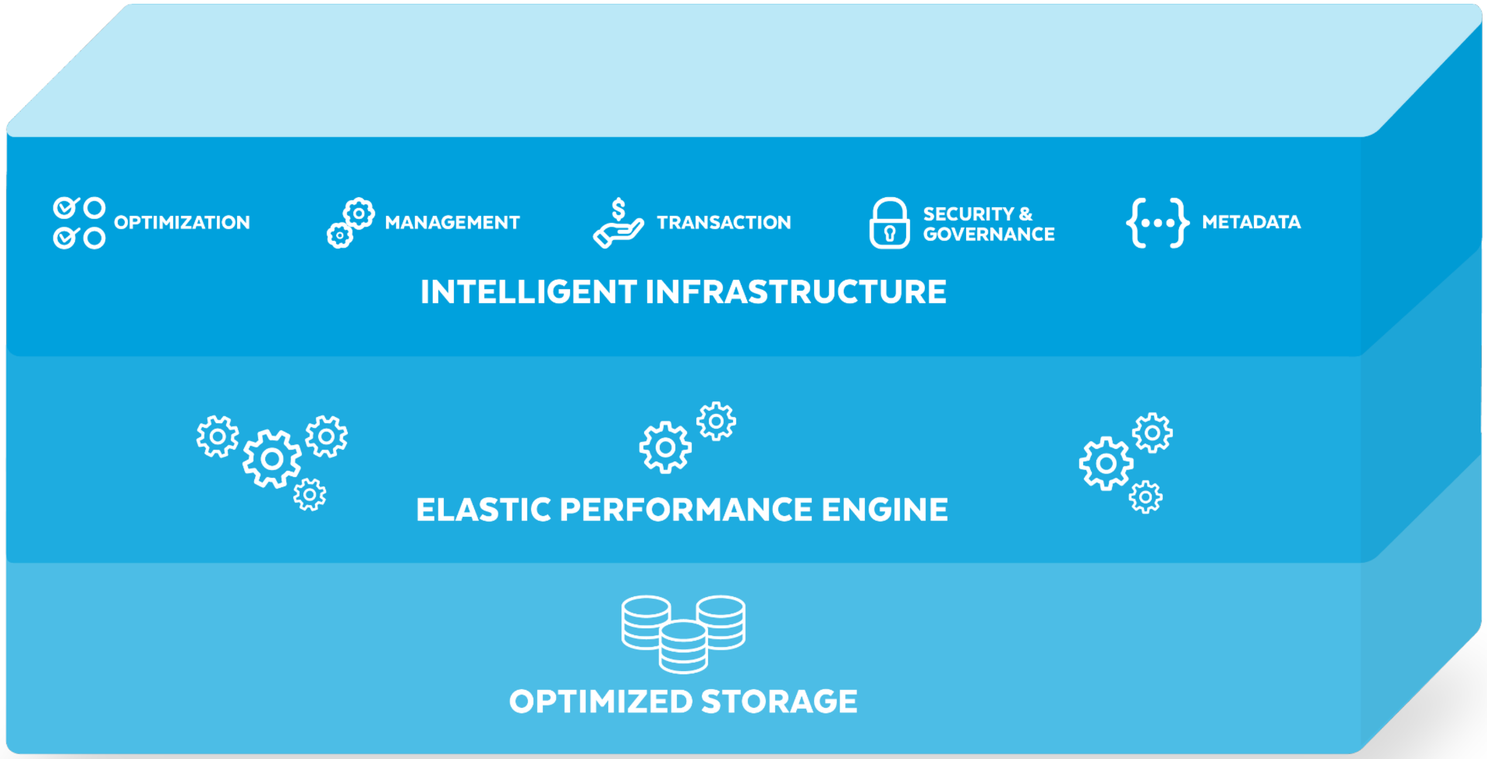
Administrar un almacén de datos de arquitectura tradicional con un rendimiento inferior al esperado

- El equipo de TI de CEMEX tenía **problemas para administrar el almacén de datos de arquitectura tradicional instalado** en sus servidores.

- Se requería la **presencia de un equipo especializado** en cada una de las regiones donde opera la empresa (México, Estados Unidos, Centroamérica, Sudamérica, el Caribe, Europa, Asia, Medio Oriente y África) **para administrar la infraestructura del almacén de datos.**

- **La administración del mantenimiento**, las actualizaciones del panel, las solicitudes de informes y la elaboración de informes mensuales **requerían demasiados recursos**, los cuales se podían aprovechar mejor en otras iniciativas más estratégicas.
- Al final de cada mes, **las actividades y los informes simultáneos generaban cuellos de botella d**e rendimiento

Plataforma Snowflake



La Solución

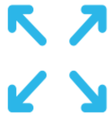
Una plataforma de datos, la cual mejora el rendimiento y reduce la carga de soporte

- CEMEX utiliza **la plataforma de Snowflake como lago de datos y como almacén de datos**, y ha creado 38 modelos con los datos de sus clientes y sus operaciones de campo.
- La plataforma de Snowflake le permite a CEMEX **almacenar datos estructurados y semiestructurados**.
- CEMEX estrenó una plataforma digital, llamada **CEMEX Go**, la cual permite automatizar los flujos de trabajo desde el pedido hasta el pago, permite realizar compras en línea y hace un seguimiento de los pedidos en tiempo real.
- Aprovechar al máximo sus capacidades de **elaboración de informes, tableros para los clientes y análisis avanzado.**
- CEMEX Go está disponible en **21 países, y casi el 90 % de los clientes recurrentes** de CEMEX utilizan esta plataforma.
- Cada año, más de **500 000 pagos y 2,5 millones de entregas** se completan por medio de CEMEX Go.

Valor de Negocio

Al contar con una plataforma de datos creada para la nube, el equipo de TI de CEMEX se puede enfocar en las iniciativas de negocio más estratégicas

Con la capacidad de almacenamiento ilimitado, CEMEX no tiene que planear mejoras en su infraestructura



Los recursos de cómputo flexibles se pueden escalar para cubrir las necesidades a corto plazo de una manera económica

Con el modelo de cobro flexible, CEMEX paga únicamente los recursos de cómputo que consume

El soporte para el almacenamiento de datos estructurados y sin estructura ofrece una fuente única de información veraz

El rendimiento es mucho mejor en comparación con las soluciones tradicionales de almacén de datos, lo que reduce la competencia por los recursos

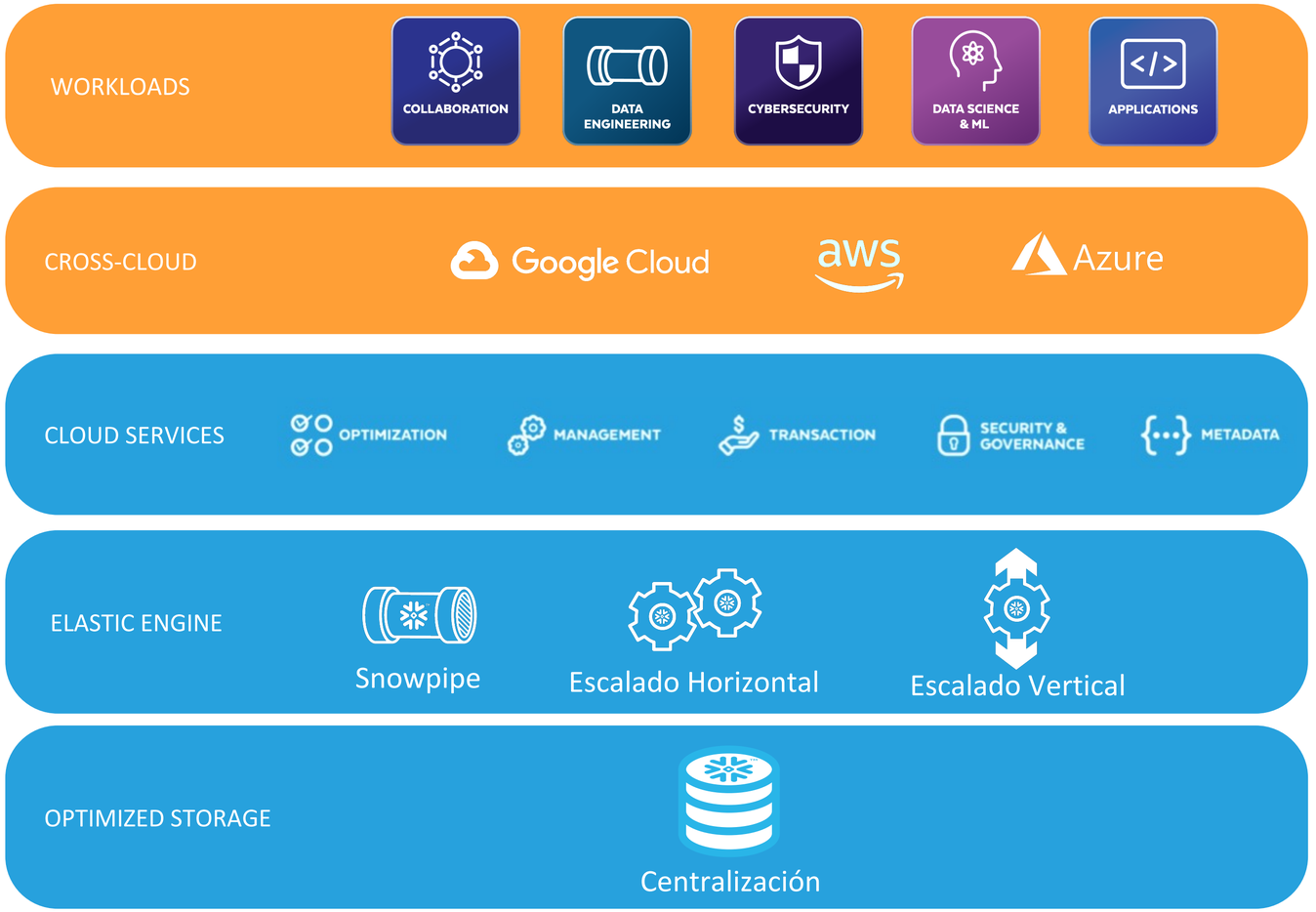
El futuro

Creación de aplicaciones ML con los datos almacenados en el lago de datos

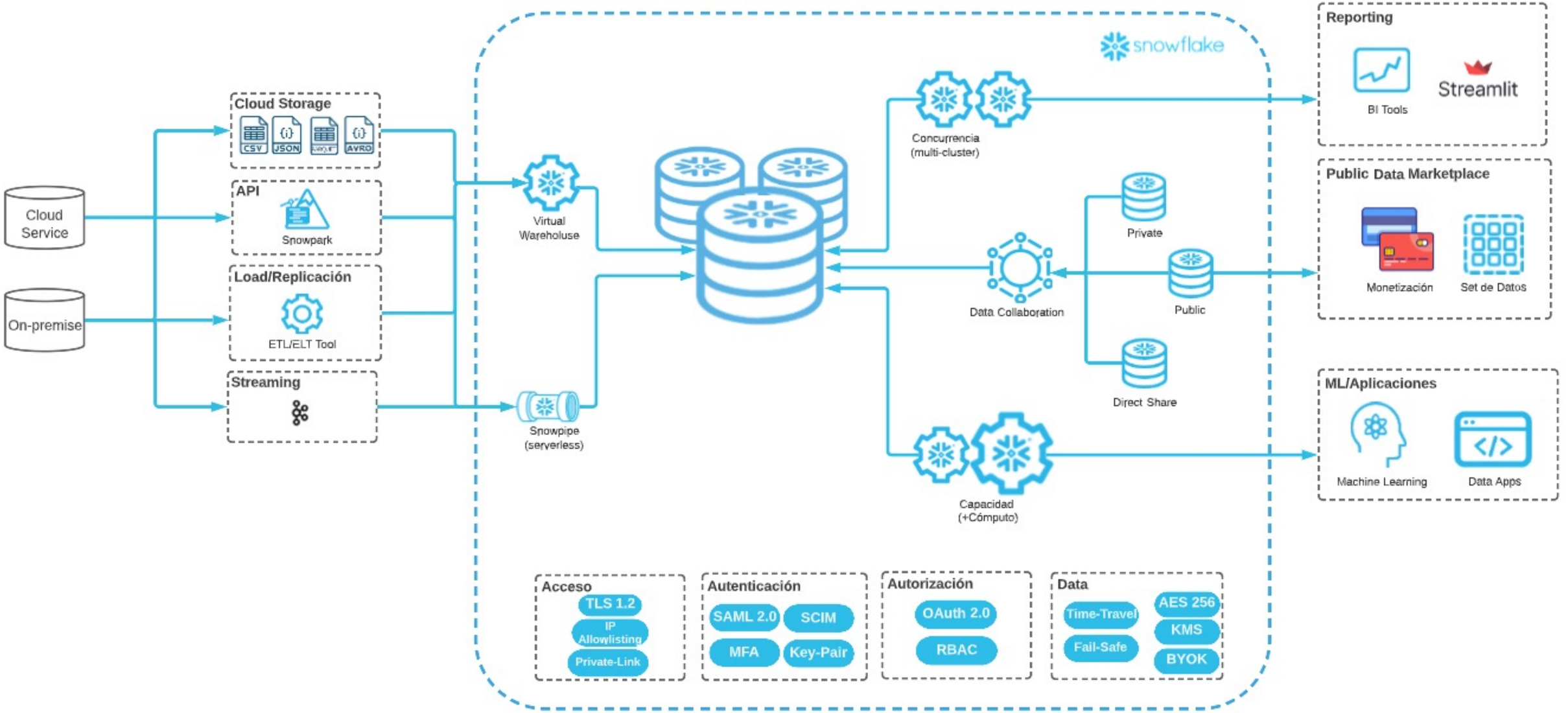
- CEMEX ya **utiliza algunas aplicaciones de ML** y planea usar más en el futuro.
- Actualmente, evalúa los **datos de tránsito y GPS almacenados en Snowflake** a fin de elegir las mejores rutas para las revolvedoras de concreto premezclado de la empresa.
- Una **aplicación similar calcula la distribución óptima de las revolvedoras** según la ubicación de las plantas de concreto premezclado de la empresa y el pronóstico de la demanda.
- En el futuro, CEMEX planea **utilizar modelos de ML en los datos de sus clientes para identificar oportunidades de ventas adicionales y ventas cruzadas**, además de ofrecer recomendaciones sobre diferentes estrategias de precios, incluidos los precios dinámicos.

snowflake®

Plataforma Snowflake



Arquitectura de Referencia



CONTEXTO

Data y Censo

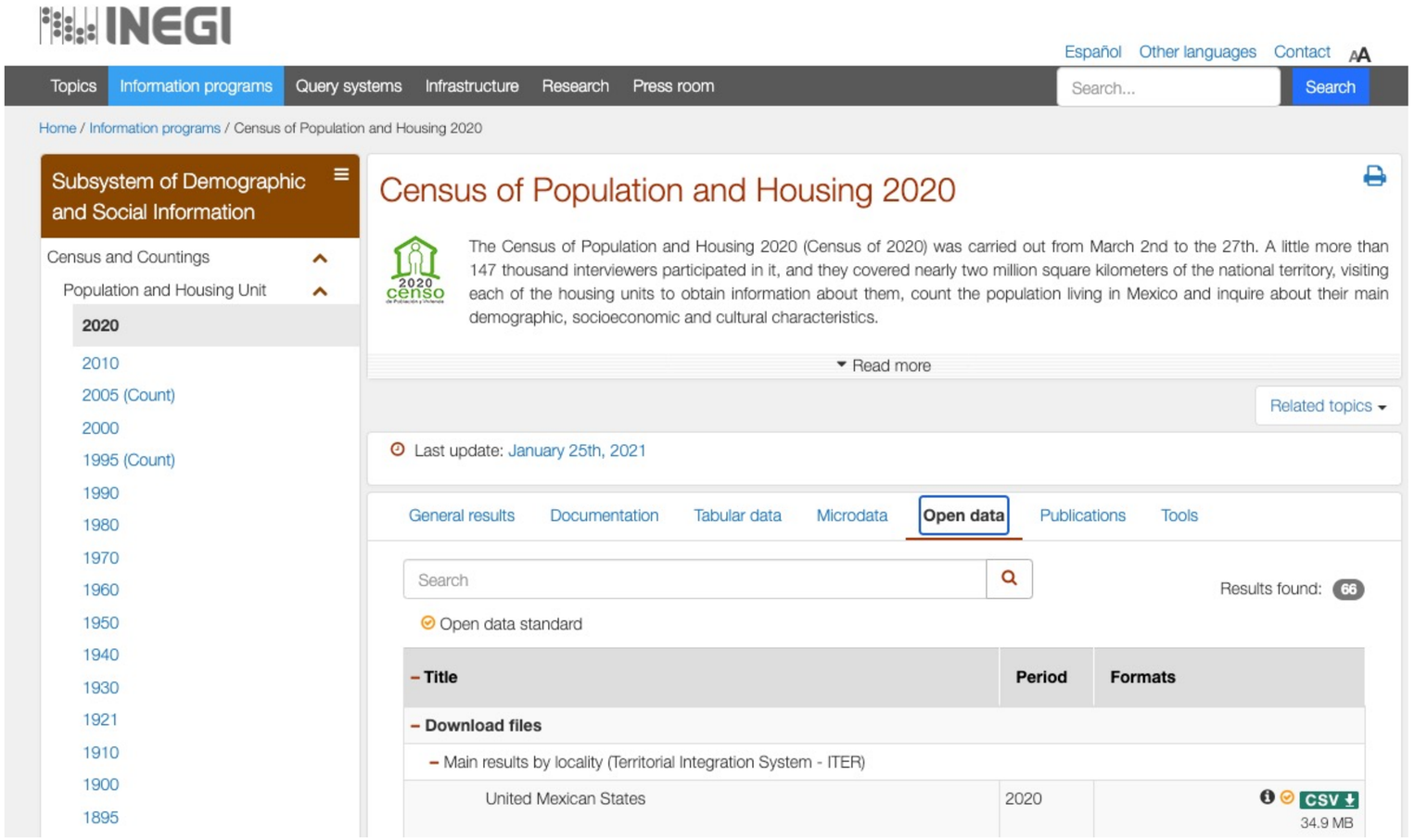
Instituto Nacional de Estadística y Geografía

- Censo Mexicano de Población y Vivienda
- Censos Económicos Mexicanos

PROBLEMA

Iniciativa de Datos Abiertos

Enfoque de autodescarga de archivos



- Alojado en un host público web no catalogado
- Limitación o compatibilidad del formato de archivo
- Referencia asíncrona de datos
- Datos no curados
- API REST limitada

https://en.www.inegi.org.mx/programas/ccpv/2020/#Open_data

Datos con tipo incorrecto o NULL

INEGI

Español Other languages Contact AA

ENTIDAD	NOM_ENT	MUN	NOM_MUN	LOC	NOM_LOC	LONGITUD	LATITUD	ALTITUD	POBTOT	POBFEM	POBMAS	P_0A2	P_0A2_F	P_0A2_M	P_3YMAS	P_3YMAS_F	P_3YMA
0	Total nacional	0	Total nacional	0	Total nacional				126014024	64540634	61473390	5764054	2848875	2915179	119976584	61554567	58422C
0	Total nacional	0	Total nacional	9998	Localidades de una vivienda				250354	96869	153485	10493	5193	5300	239441	91463	14797
0	Total nacional	0	Total nacional	9999	Localidades de dos viviendas				147125	61324	85801	6798	3407	3391	139757	57628	82129
1	Aguascalientes	0	Total de la entidad Aguascalientes	0	Total de la Entidad				1425607	728924	696683	71864	35604	36260	1352235	692561	65967
1	Aguascalientes	0	Total de la entidad Aguascalientes	9998	Localidades de una vivienda				3697	1510	2187	165	81	84	3532	1429	2103
1	Aguascalientes	0	Total de la entidad Aguascalientes	9999	Localidades de dos viviendas				3021	1013	2008	119	54	65	2902	959	1943
1	Aguascalientes	1	Aguascalientes	0	Total del Municipio				948990	486917	462073	44372	21893	22479	903684	464556	43912
1	Aguascalientes	1	Aguascalientes	1	Aguascalientes	102°17'45.768" W	21°52'47.362" N	1878	863893	444725	419168	39525	19552	19973	823490	424733	39875
1	Aguascalientes	1	Aguascalientes	94	Granja Adelita	102°22'24.710" W	21°52'18.749" N	1902	5	*	*	*	*	*	*	*	*
1	Aguascalientes	1	Aguascalientes	96	Agua Azul	102°21'25.639" W	21°53'01.522" N	1861	41	17	24	2	2	0	39	15	24
1	Aguascalientes	1	Aguascalientes	102	Los Arbolitos [Rancho]	102°21'26.261" W	21°46'48.650" N	1861	8	*	*	*	*	*	*	*	*
1	Aguascalientes	1	Aguascalientes	104	Ardillas de Abajo (Las Ardillas)	102°11'30.914" W	21°56'42.243" N	1989	1	*	*	*	*	*	*	*	*
1	Aguascalientes	1	Aguascalientes	106	Arellano	102°16'26.238" W	21°48'06.384" N	1892	1169	556	613	53	25	28	1116	531	585
1	Aguascalientes	1	Aguascalientes	112	Bajío los Vázquez	102°07'29.341" W	21°44'50.978" N	1971	41	21	20	1	0	1	40	21	19
19	Nuevo León	38	Montemorelos	559	Septiembre	99°52'03.379" W	25°11'39.299" N	430	31	16	15	1	0	1	30	16	14
19	Nuevo León	38	Montemorelos	562	Tierras Coloradas	99°52'46.403" W	25°16'39.292" N	341	85	43	42	4	3	1	81	40	41
19	Nuevo León	38	Montemorelos	564	Los Vaquero (La Cuesta)	99°54'12.025" W	25°12'03.705" N	437	3	*	*	*	*	*	*	*	*
19	Nuevo León	38	Montemorelos	565	Loma Prieta	99°59'29.737" W	25°15'38.151" N	441	143	64	79	8	0	8	135	64	71
19	Nuevo León	38	Montemorelos	566	San Pedro de la Rosa	99°48'49.471" W	25°05'01.287" N	498	9	5	4	0	0	0	9	5	4
19	Nuevo León	38	Montemorelos	567	San Jorge	99°42'21.161" W	25°03'58.985" N	405	7	*	*	*	*	*	*	*	*
19	Nuevo León	38	Montemorelos	570	Nuprim	99°44'49.294" W	25°00'23.041" N	492	4	*	*	*	*	*	*	*	*
19	Nuevo León	38	Montemorelos	571	San Juan de Ocampo	99°51'44.275" W	25°17'33.628" N	320	9	3	6	0	0	0	9	3	6

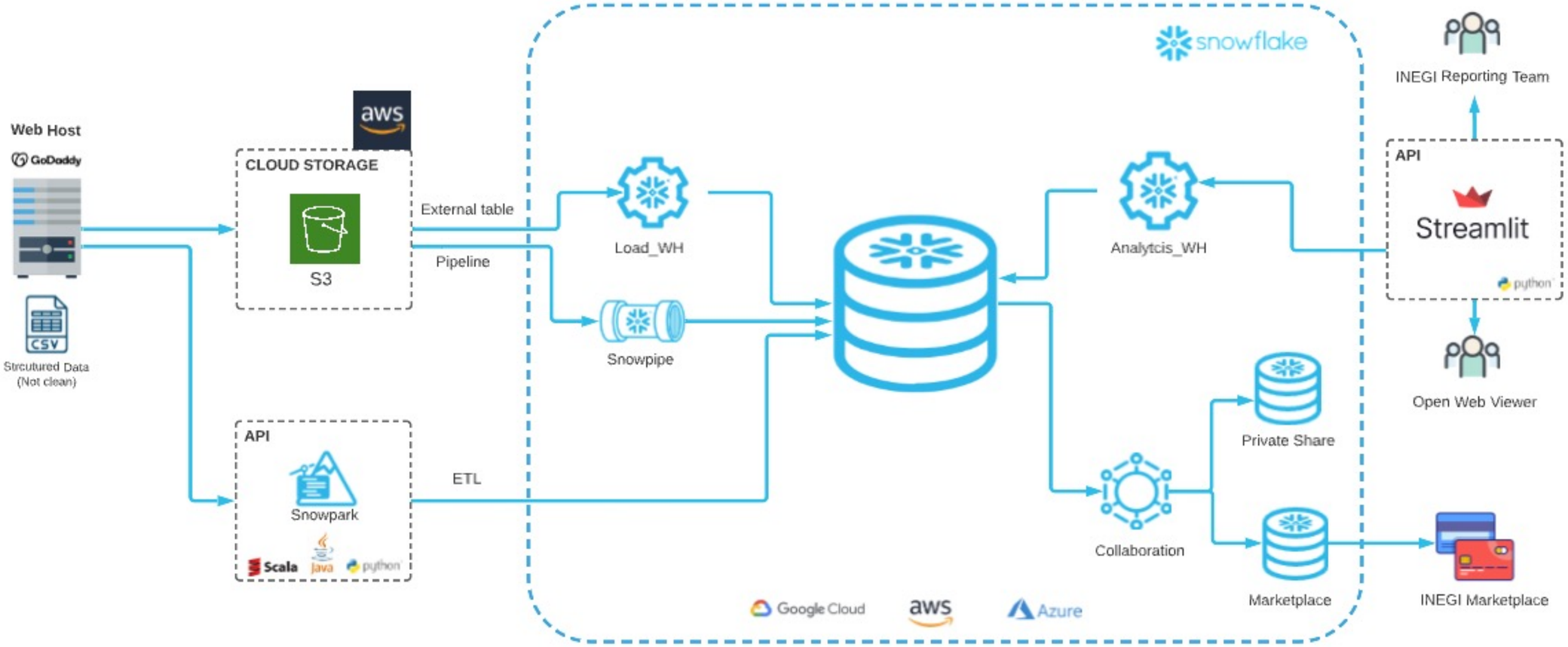
DISEÑO DE SOLUCIÓN

+

Snowpark API

+

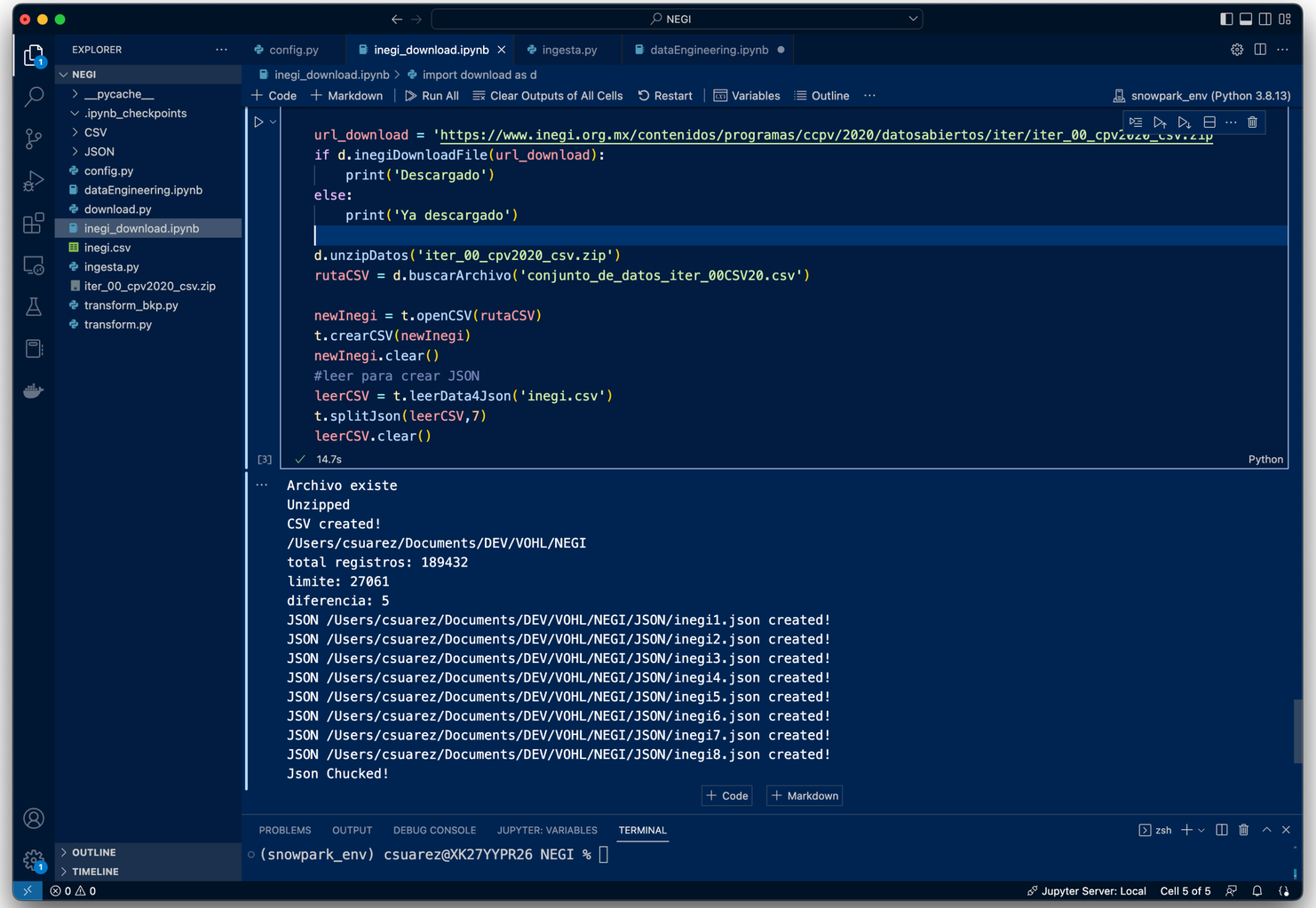
National Institute of Statistics and Geography



https://en.www.inegi.org.mx/datosabiertos/

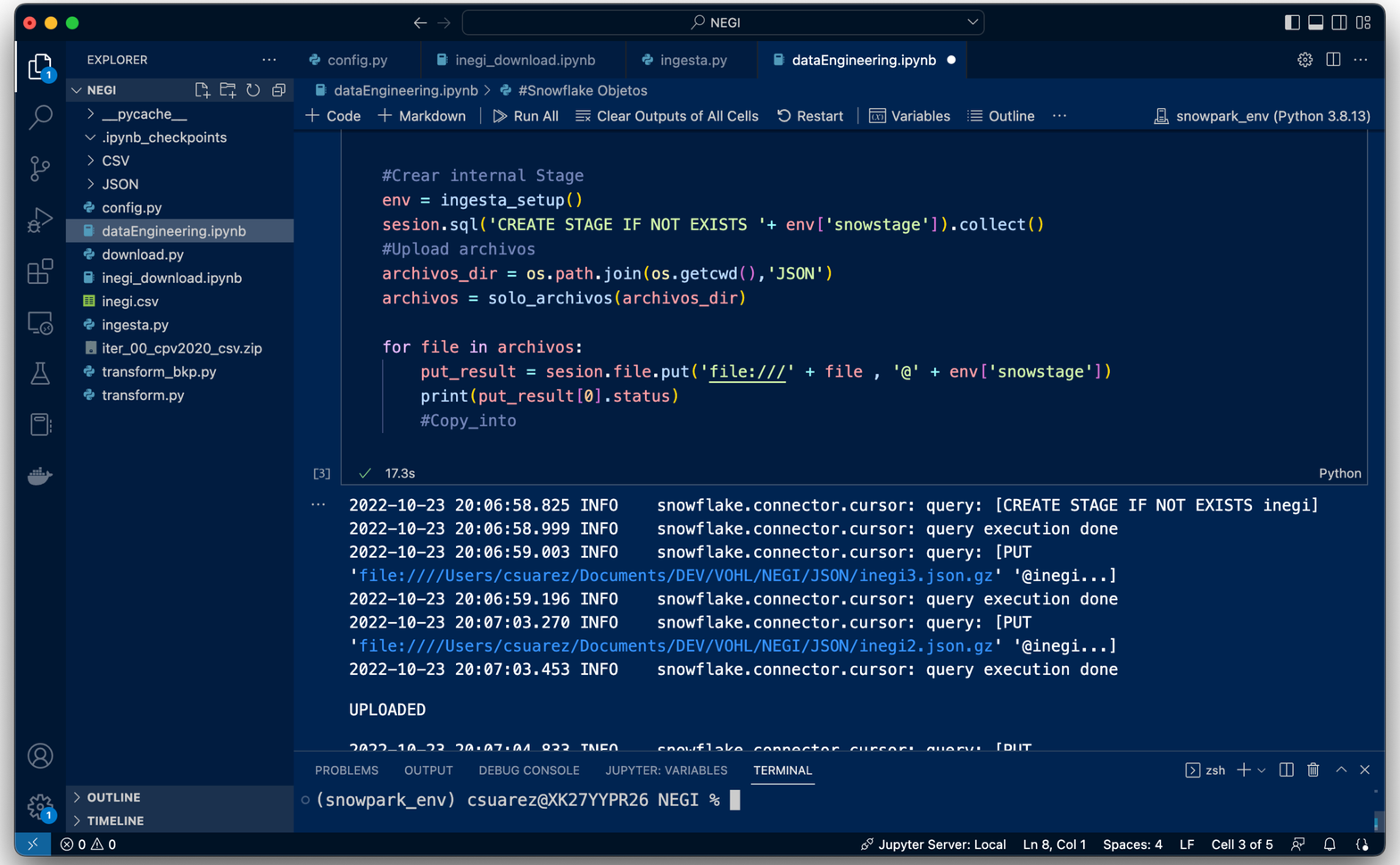
CODING

Extraer y transformar Python



- Descargar archivos de Web Host
- Transformar (NULL, Grados, números)
- Serializar a JSON
- Dividir JSON en micro-batch

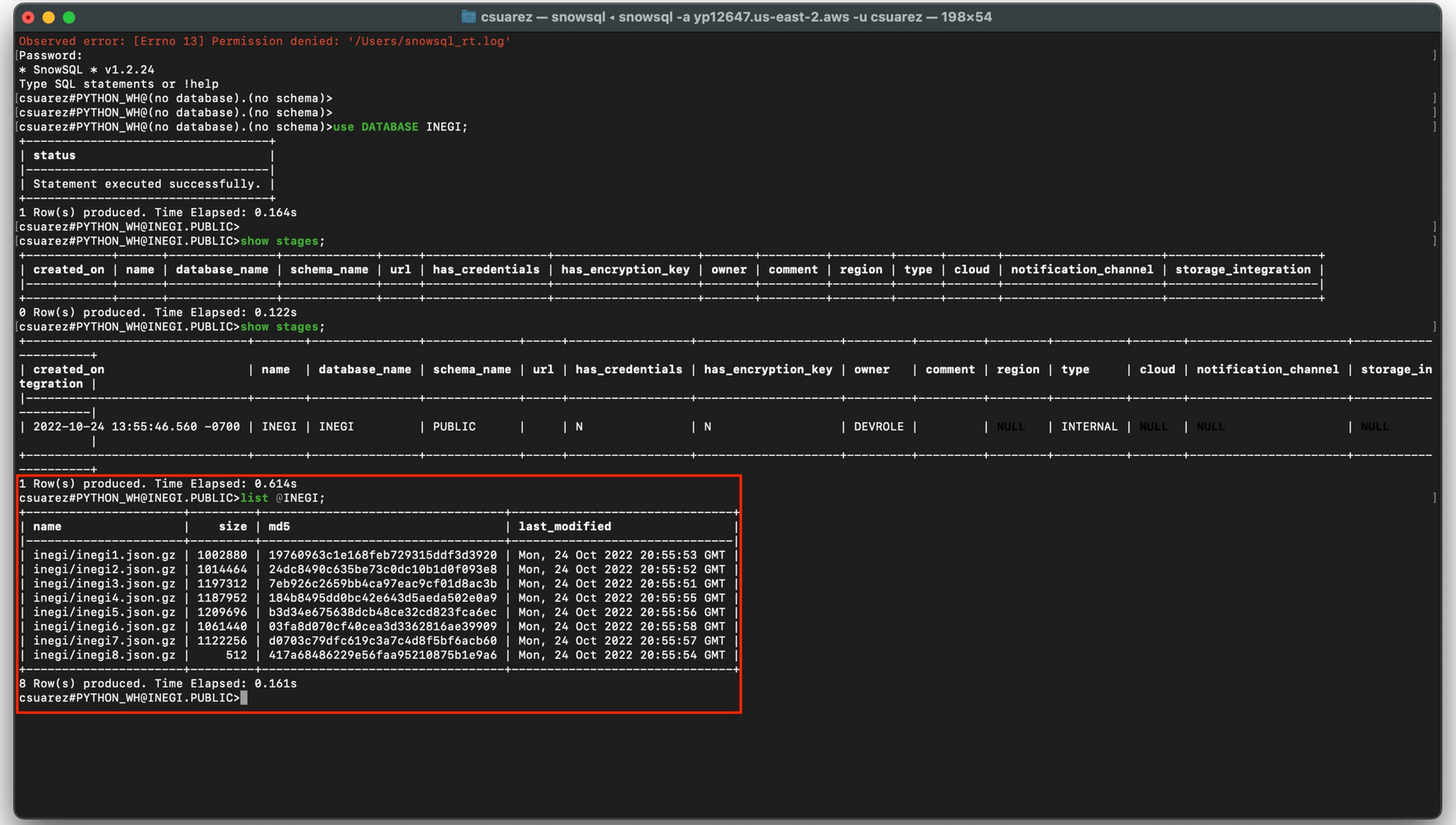
Python - Ingeniería y Carga de Datos



Snowpark API

- Session object
- Virtual warehouse
- Snowpark API (Stages, PUT)

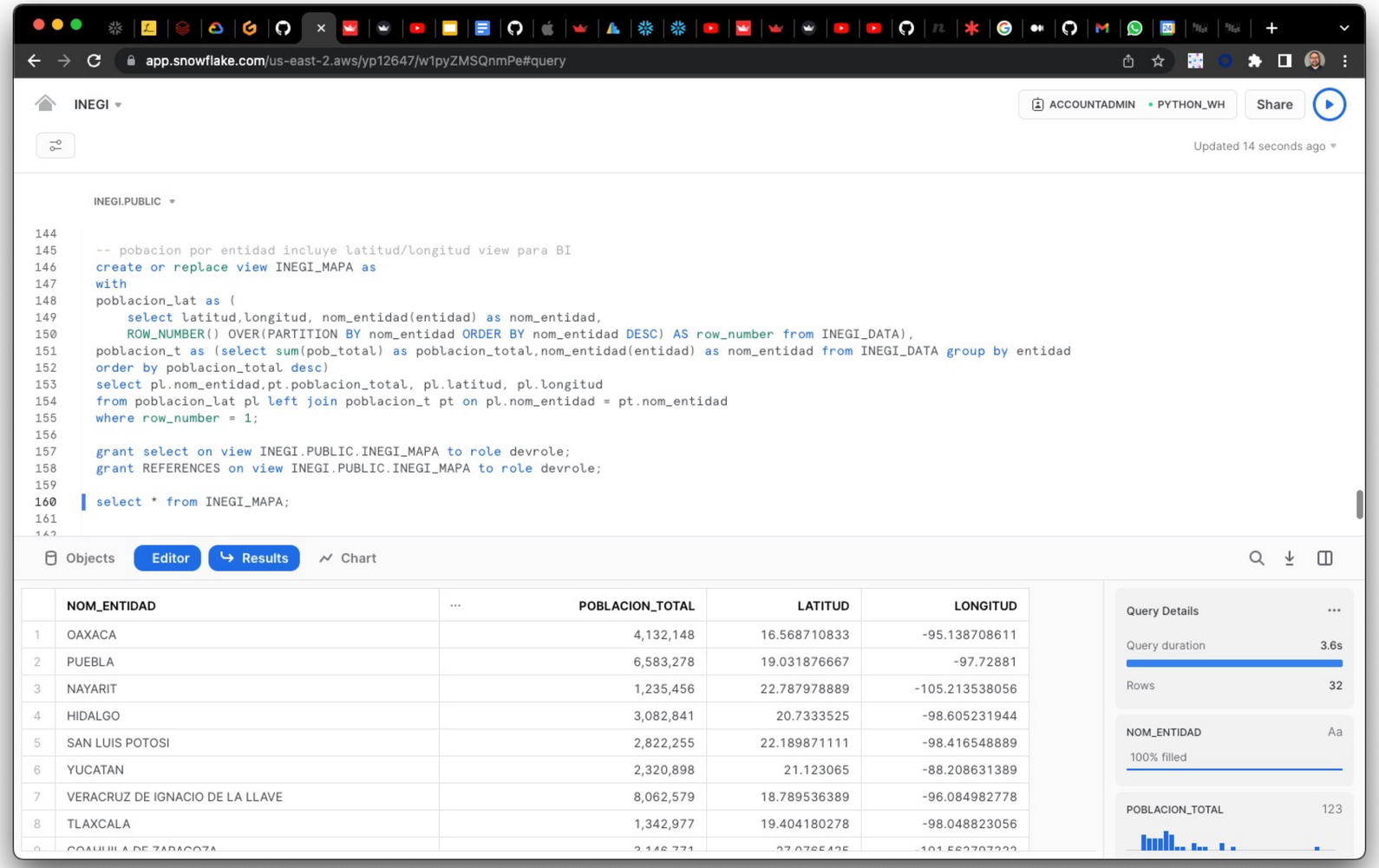
Validación Carga de Data



SnowSQL

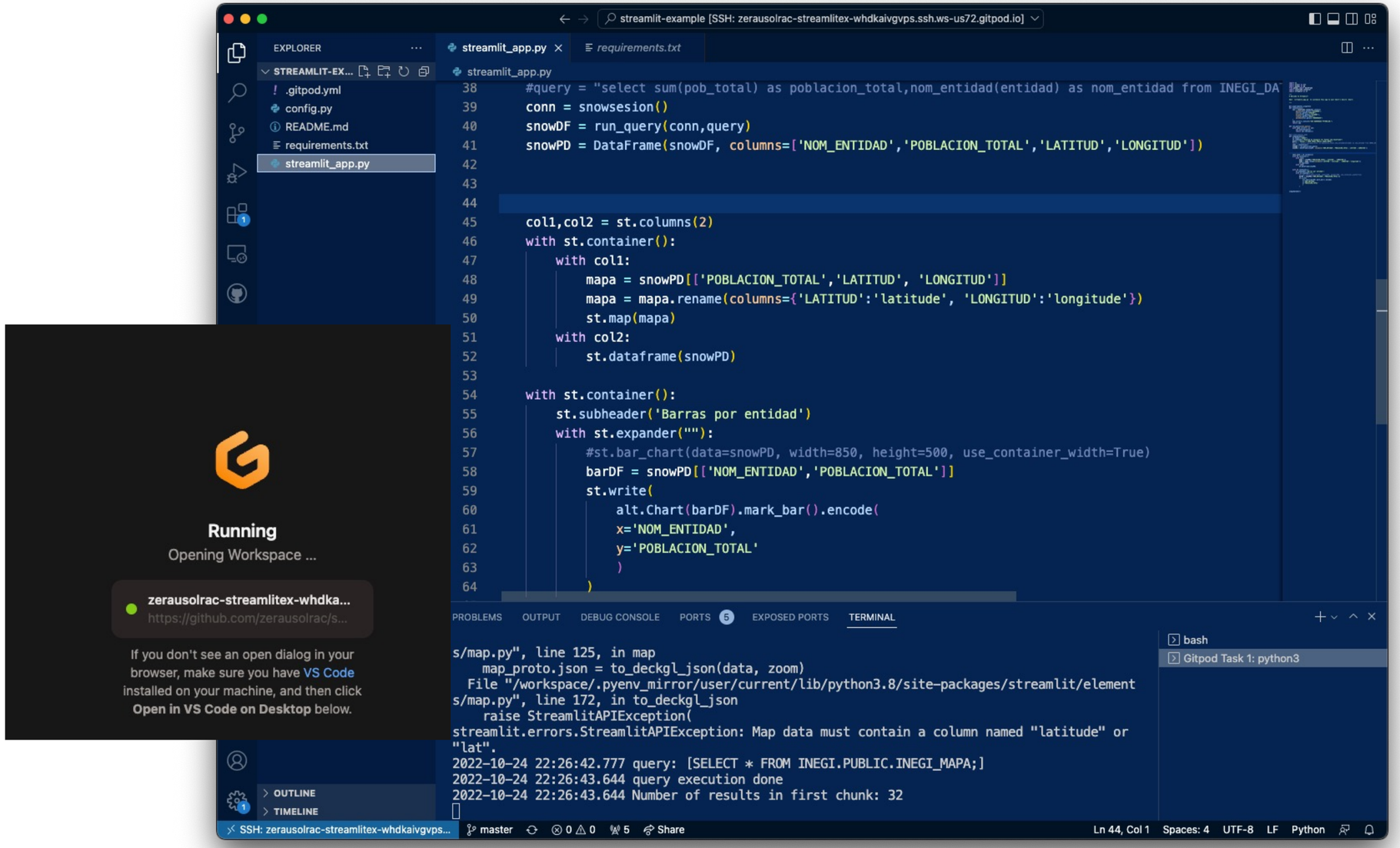
- Valide que el archivo JSON exista en el almacenamiento interno

Data Model y Views sobre los Datos



- Python UDF
- Privileges and roles
- Datalake RAW data
- Data View

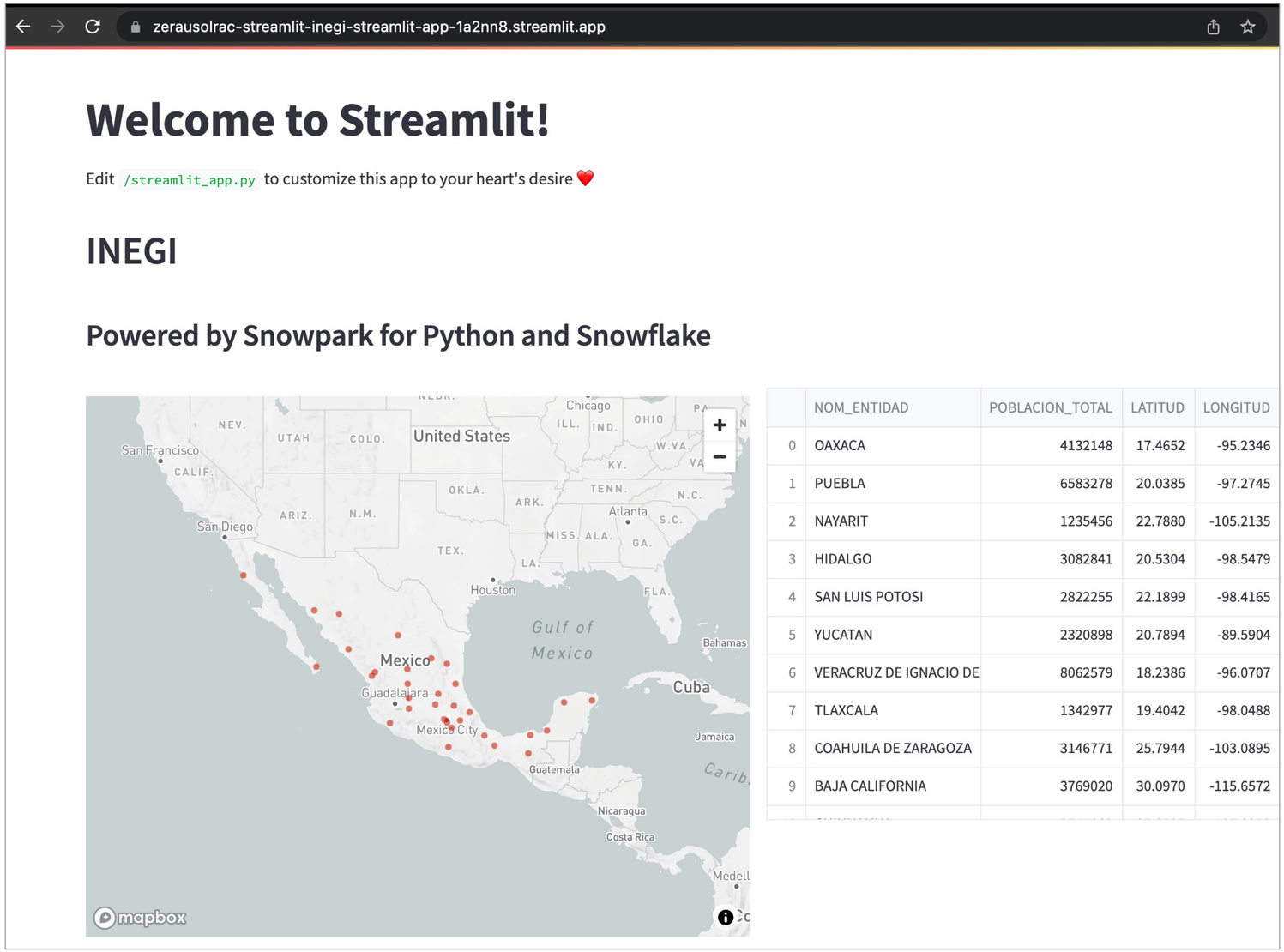
Coding Streamlit / Cloud Repo



Streamlit

- Lectura desde Snowflake (Snowpark)
- Configuración variables env (encriptadas)
- Acomodo con Streamlit containers y datos sobre Streamlit vars

Visualización de Datos - INEGI



	NOM_ENTIDAD	POBLACION_TOTAL	LATITUD	LONGITUD
0	OAXACA	4132148	17.4652	-95.2346
1	PUEBLA	6583278	20.0385	-97.2745
2	NAYARIT	1235456	22.7880	-105.2135
3	HIDALGO	3082841	20.5304	-98.5479
4	SAN LUIS POTOSI	2822255	22.1899	-98.4165
5	YUCATAN	2320898	20.7894	-89.5904
6	VERACRUZ DE IGNACIO DE	8062579	18.2386	-96.0707
7	TLAXCALA	1342977	19.4042	-98.0488
8	COAHUILA DE ZARAGOZA	3146771	25.7944	-103.0895
9	BAJA CALIFORNIA	3769020	30.0970	-115.6572

Streamlit

https://zerausolrac-streamlit-inegi-streamlit-app-1a2nn8.streamlit.app/

¡GRACIAS!

Carlos Suárez
carlos.suarez@snowflake.com