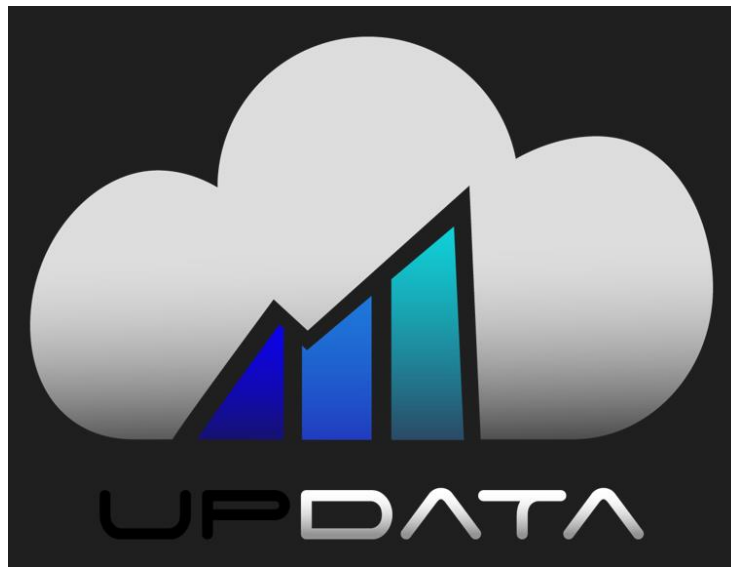


**PROYECTO:**

**RESTAURANTES EN EL ESTADO DE FLORIDA,  
ESTADOS UNIDOS**



**DOCUMENTACIÓN ETAPA DE INGENIERÍA DE DATOS**

Ornaldo Hernández Ramos  
Cristian Ignacio Papini  
Aaron Bernardo Martínez Vera  
Jhoeliel David Palma Salazar  
Janice Rocío Rico Sánchez

## DATASETS Y DICCIONARIO DE DATOS

Los Datasets suministrados están divididos en varias carpetas con la siguiente distribución:

<b>Carpeta</b>	<b>Sub-Carpeta</b>	<b>Cantidad Archivos .json</b>	<b>Peso (Mb)</b>
reviews-estados	review-Alabama	12	467,90
	review-Alaska	4	142,30
	review-Arizona	14	636,00
	review-Arkansas	16	611,80
	review-California	18	762,90
	review-Colorado	16	751,50
	review-Connecticut	18	673,60
	review-Delaware	7	235,00
	review-District_of_Columbia	4	160,90
	review-Florida	19	886,30
	review-Georgia	13	575,50
	review-Hawaii	11	471,70
	review-Idaho	14	569,50
	review-Illinois	14	587,50
	review-Indiana	15	599,00
	review-Iowa	18	673,00
	review-Kansas	13	511,00
	review-Kentucky	11	429,50
	review-Louisiana	10	385,30
	review-Maine	8	290,00
	review-Maryland	16	669,30
	review-Massachusetts	16	640,60
	review-Michigan	15	624,70
	review-Minnesota	12	502,80
	review-Mississippi	14	486,70
	review-Missouri	11	455,90
	review-Montana	7	248,00
	review-Nebraska	13	468,30
	review-Nevada	12	520,00
	review-New_Hampshire	9	342,40
	review-New_Jersey	13	539,80
	review-New_Mexico	12	486,90
	review-New_York	18	759,90
	review-North_Carolina	15	656,20
	review-North_Dakota	4	144,90
	review-Ohio	13	538,50
	review-Oklahoma	11	445,80
	review-Oregon	15	644,90
	review-Pennsylvania	16	658,30
	review-Rhode_Island	6	227,20
	review-South_Carolina	14	575,60
	review-South_Dakota	5	172,60

	review-Tennessee	12	497,40
	review-Texas	16	713,50
	review-Utah	10	477,20
	review-Vermont	3	85,60
	review-Virginia	12	480,80
	review-Washington	13	569,80
	review-West_Virginia	8	264,80
	review-Wisconsin	12	454,10
	review-Wyoming	3	111,00
	<b>Total Archivos</b>	<b>611</b>	
		<b>Total Peso (Mb)</b>	<b>24.883,70</b>
metadata-sitios		<b>11</b>	<b>2.832,30</b>

Cada archivo del Dataset **‘reviews-estados’** contiene los siguientes campos:

Columna	Detalle	Tipo de Dato	Ejemplo
user_id	Id de usuario	float64	1.00510015423490261984
name	nombre	object	Ashley Cassidy
time	hora	int64	1597292102991
rating	puntuación	int64	5
text	comentario	object	Love this place, and Katie B. is an artist with color!
pics	fotos	object	[{"url": ["https://lh5.googleusercontent.com/p/AF1QipPDBZQ6hnEg8eZOITUO9NenSWqCPXYjs t55UK3P=w150-h150-k-no-p"]}]]
resp	respuesta	object	{"time": 1618596955963, "text": "Thank you, Ashley!"}
gmap_id	Id de gmap	object	0x886267a5d761a8e3:0xda004ea8eefc3e37

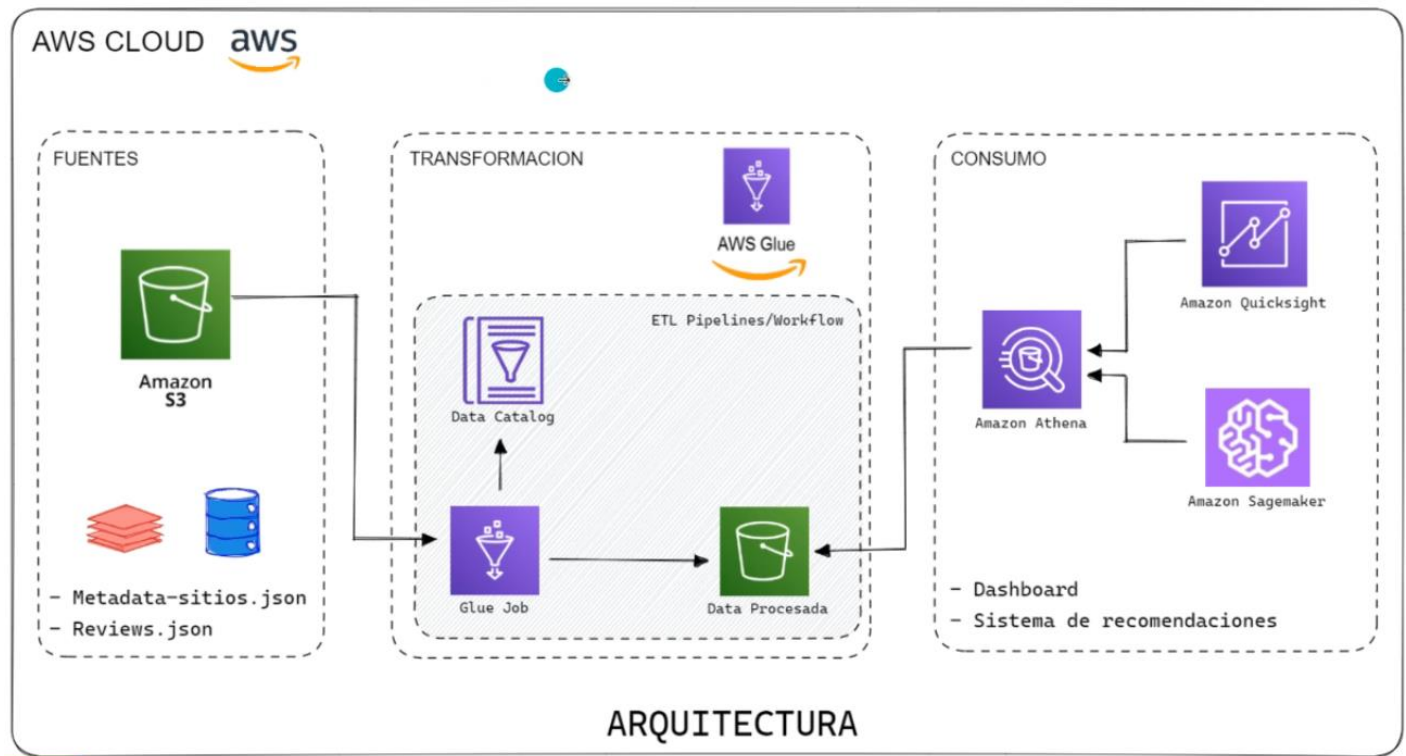
Cada archivo del Dataset **‘metadata-sitios’** contiene los siguientes campos:

Columna	Detalle	Tipo de Dato	Ejemplo
name	nombre	object	Redbox
address	dirección	object	Redbox, 2214 Muscatine Ave, Iowa City, IA 52240
gmap_id	Id de gmap	object	0x87e46a6a9334aeb3:0xe1054df61eb304e
description	descripción	object	Kiosk chain renting movies, including new releases (many are located inside stores).
latitude	latitud	float64	41.649925599999996
longitude	longitud	float64	-91.502687
category	categoría	object	["Movie rental kiosk", "DVD store", "Movie rental store"]
avg_rating	promedio puntuación	float64	5
num_of_reviews	cantidad de reseñas	int64	4
price	precio	object	\$

hours	horarios	object	[[{"Saturday", "Open 24 hours"}, {"Sunday", "Open 24 hours"}, {"Monday", "Open 24 hours"}, {"Tuesday", "Open 24 hours"}, {"Wednesday", "Open 24 hours"}, {"Thursday", "Open 24 hours"}, {"Friday", "Open 24 hours"}]]
MISC	misceláneos	object	{"Accessibility": ["Wheelchair accessible entrance"]}
state	estado del lugar	object	Open 24 hours
relative_results	resultados relativos	object	["0x87e441e7d8c5cd21:0x4433ec9be70020c8", "0x87e46a70197a6217:0x51fb6854769d020b", "0x87e443f0e024a085:0x3647481adf689deb", "0x87e4416b12de92f9:0x4e88f99fdec82927"]
url	web en Google Maps	object	<a href="https://www.google.com/maps/place//data=!4m2!3m1!1s0x87e46a6a9334aeb3:0xe1054df61eb304e?authuser=-1&amp;hl=en&amp;gl=us">https://www.google.com/maps/place//data=!4m2!3m1!1s0x87e46a6a9334aeb3:0xe1054df61eb304e?authuser=-1&amp;hl=en&amp;gl=us</a>

# WORKFLOW DE TECNOLOGÍAS

## DATA WAREHOUSE



## DIAGRAMA ER

La relación que se encontró entre las dos tablas (***metadata-sitios*** y ***reviews-estados***) es a través del campo *relative\_results* de la primera y *gmap\_id* de la segunda, lo que será conveniente para los procesos de Análisis de Datos que seguirán posteriormente.

