

## 1. Obtenemos la geolocalización de NewYork (EEUU)

Dirección

[Obtener Coordenadas GPS](#)

GD (grados decimales)\*

Latitud

Longitud

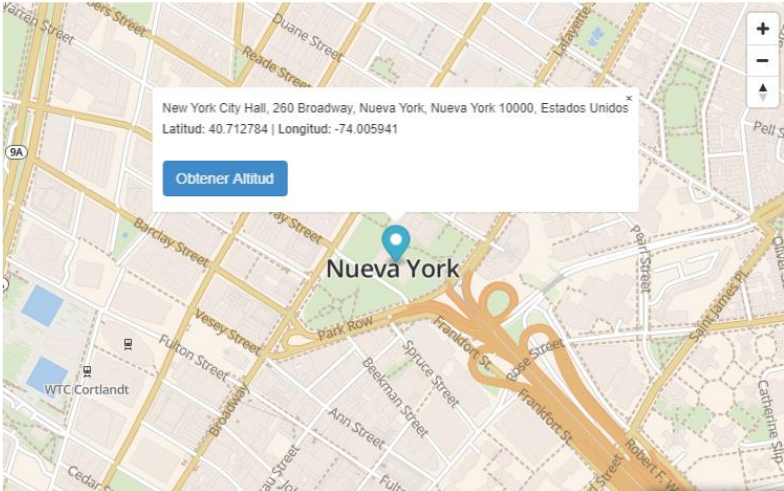
[Obtener Dirección](#)

GMS (grados, minutos, segundos)\*

Latitud ☐ N ☐ S ° ′ "

Longitud ☐ E ☐ O ° ′ "

[Obtener Dirección](#)



## 2. Obtenemos las credenciales de nuestro api en twitter

### Keys and tokens

### Permissions

#### Keys and tokens

Keys, secret keys and access tokens management.

#### Consumer API keys

[Regenerate](#)

**API key:** URgKLmIHaoUGggjynVvgELYqz

**API secret key:** dLnT08kNMPB5hzEPjlsIPdYtw2KAVtkAzujXAnjFX016EvJSTR

#### Access token & access token secret

[Revoke](#)

[Regenerate](#)

We only show your access token and secret when you first generate it in order to make your account more secure. You can revoke or regenerate them at any time, which will invalidate your existing tokens.

**Access token:** xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Last generated: Jul 8, 2020

3. Creamos y ejecutamos el script `tweets_newyork.py`

[illegible]

#### 4. Verificamos nuestra base en couchdb

Databases

Database name

Create Database

dbcanciones	74.2 KB	17	No			
dblibros	36.5 KB	8	No			
destino	2.0 MB	748	No			
destino_delitos	54.9 KB	114	No			
ejemplo	102.1 KB	43	No			
ejemplo4	12.9 MB	4755	No			
guayaquil	9.6 KB	8	No			
gye	8.9 KB	8	No			
newyork	0.9 MB	387	No			
pedidos	9.4 KB	11	No			

Showing 1 of 10 databases




























Databases overview

20





5. Replicamos los pasos anteriores creando el scrip `tweets_covid.py` y lo ejecutamos

[illegible]

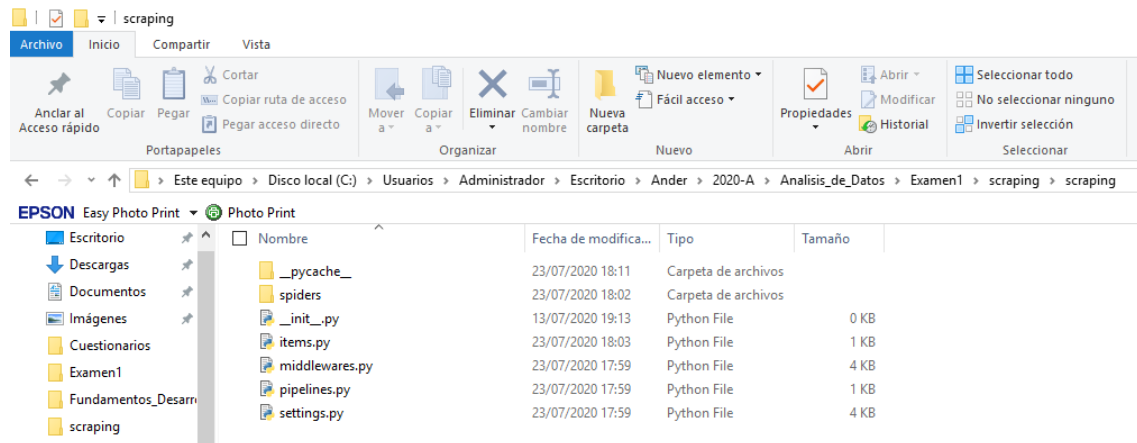
## 6. Verificamos nuestra base en cocuchdb

Databases				
Database name		Create Database		
Name	Size	# of Docs	Partitioned	Actions
_replicator	33.0 KB	16	No	  
agenda	2.3 KB	11	No	  
ambato	10.1 KB	8	No	  
canada	2.7 KB	1	No	  
canasta_compras	12.8 KB	6	No	  
chente	10.4 MB	4132	No	  
covid	0.7 MB	264	No	  
cuenca	10.7 KB	8	No	  
db1	3.3 MB	1130	No	  

## 7. Creamos una replica desde la base de newyork hacia la base covid

	<a href="http://localhost:5984/newyork">http://localhost:5984/newyork</a>	<a href="http://localhost:5984/covid">http://localhost:5984/covid</a>	Jul 23rd, 7:32 pm	Continuous	Running			
---	---	---	-------------------	------------	---------	---	---	---

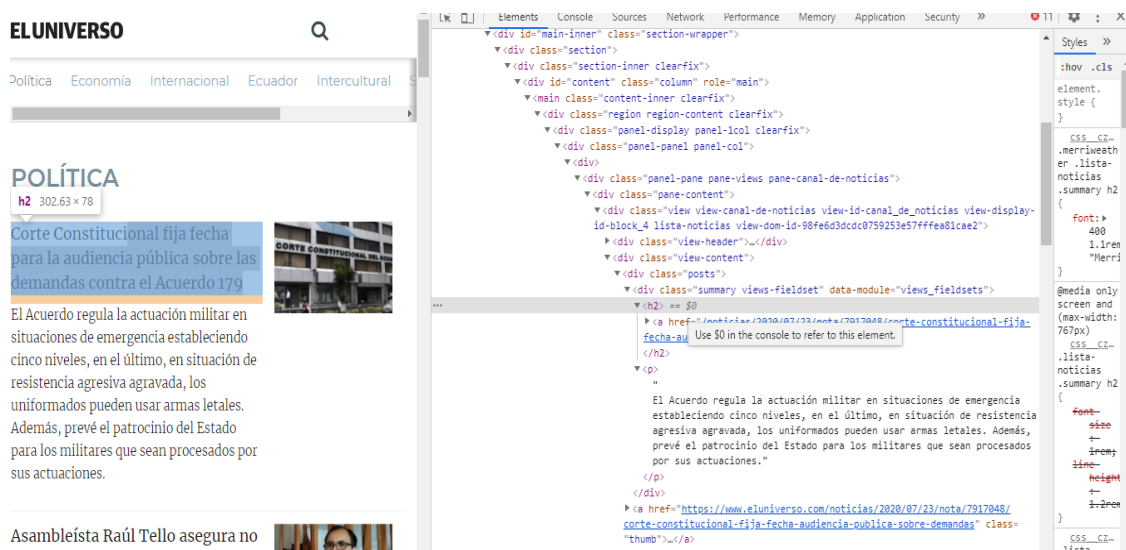
## 8. Creamos un nuevo project scrapy



## 9. Actualizamos los items.py que vamos a utilizar para nuestra búsqueda

```
1 # Define here the models for your scraped items
2 #
3 # See documentation in:
4 # https://docs.scrapy.org/en/latest/topics/items.html
5
6 import scrapy
7 from scrapy.item import Item, Field
8
9
10 class ScrapingItem(scrapy.Item):
11     # define the fields for your item here like:
12     # name = scrapy.Field()
13     title=Field()
14     url=Field()
15     pass
16
```

## 10. En la carpeta spiders creamos el script de universo\_spiders.py. Examinando la estructura de la pagina web del universo



## 11. En la ubicación de la carpeta scraping corremos el comando crawl para recuperar los datos en un json

```
Administrador: C:\Windows\System32\cmd.exe - scrapy crawl news_universo -o news.json -t json
{'title': 'Asambleístas rinden versión en caso por delincuencia organizada ',
 'url': 'www.eluniverso.com/noticias/2020/07/23/nota/7916478/asambleistas-badi-llegan-fiscalia-rendir-version-caso-delincuencia'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/politica>
{'title': 'Cómo funciona el Programa de Protección a Víctimas y Testigos en '
 'Ecuador',
 'url': 'www.eluniverso.com/noticias/2020/07/22/nota/7915707/como-funciona-programa-proteccion-victimas-testigos-ecuador'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/politica>
{'title': 'Asambleístas rendirán versión en caso por delincuencia organizada ',
 'url': 'www.eluniverso.com/noticias/2020/07/22/nota/7915506/badi-version-delincuencia-organizada-eliseo-azuero-daniel-mendoza'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/ecuador>
{'title': 'CJN rechaza petición de suspensión condicional de la pena '
 'solicitada por Norma Vallejo',
 'url': 'www.eluniverso.com/noticias/2020/07/23/nota/7916919/tribunal-niega-suspension-condicional-pena-concusion-solicitada'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/politica>
{'title': 'Corte Constitucional dirimirá competencias de la Contraloría y del '
 'Tribunal Electoral',
 'url': 'www.eluniverso.com/noticias/2020/07/22/nota/7915659/elecciones-presidenciales-ecuador-2021-contraloria-fuerza'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/politica>
{'title': 'Organización civil apunta a una consulta popular sobre las '
 'autonomías regionales',
 'url': 'www.eluniverso.com/noticias/2020/07/22/nota/7915566/organizacion-civil-apunta-consulta-popular-sobre-autonomias'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/ecuador>
{'title': 'Luis Gallegos llama a la Celac y a China a aunar esfuerzos para que '
 'la vacuna del COVID-19 sea un 'bien público de la salud mundial'',
 'url': 'www.eluniverso.com/noticias/2020/07/23/nota/7916901/luis-gallegos-llama-celac-china-aunar-esfuerzos-que-vacuna-covid-19'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/ecuador>
{'title': 'Gobierno promueve la educación on-line para las instituciones de '
 'educación superior por la emergencia sanitaria del COVID-19',
 'url': 'www.eluniverso.com/noticias/2020/07/23/nota/7916646/gobierno-promueve-educacion-line-instituciones-educacion-superior'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/intercultural>
{'title': '"Mujer de Piedra", adaptada para teatro online, está en el Programa '
 'de Circulación de Obras Artísticas #telón2020',
 'url': 'www.eluniverso.com/entretenimiento/2020/07/09/nota/7900350/obra-mujer-piedra-presenta-casa-cultura-ecuatoriana-directo'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/internacional>
{'title': 'La poco conocida Milicia Marítima de China, la unidad de Pekín que '
 '"actúa bajo los radares" para controlar las aguas de Asia Pacífico',
 'url': 'www.eluniverso.com/noticias/2020/07/23/nota/7916880/china-milicia-maritima-pekín-radares-asia-pacifico'}
2020-07-23 18:11:54 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.eluniverso.com/ecologia>
{'title': '#SOSGalápagos, la etiqueta usada para exigir que la flota de barcos '
 'chinos se aleje del archipiélago',
 'url': 'www.eluniverso.com/noticias/2020/07/23/nota/7916694/sosgalapagos-etiqueta-flota-china-barcos-galapagos-ecuador-redes'}
2020-07-23 18:11:55 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://www.eluniverso.com/comunidad> (referer: None)
2020-07-23 18:11:55 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://www.eluniverso.com/via-costa> (referer: None)
```

## 12. Cargamos nuestro json a nuestra base de datos mongo

Default file

C:\Users\Administrador\Desktop\Ander\2020-A\Análisis\_de\_Datos\Examei [BROWSE](#)

Select Input File Type

☒ JSON ☐ CSV

Options

☒ Ignore empty strings ☐ Stop on errors

Specify Fields and Types

	title
1	Corte Constitucional fija fecha para la audiencia pública sobre las demandas contra el Ac
2	Asambleísta Raúl Tello asegura no haber conocido de esa "relación de carácter delincuenti
3	CJN rechaza petición de suspensión condicional de la pena solicitada por Norma Vallejo
4	Luis Gallegos llama a la Celac y a China a aunar esfuerzos para que la vacuna del COVID-1
5	Gabriel Manzur y Silvana V. continuarán con prisión preventiva en caso por enriquecimiento
6	En cuatro semanas saldría el informe de la investigación política a la entrega de carnés
7	La flota pesquera internacional está en el corredor de aguas internacionales, entre las z
8	Tribunal Contencioso Electoral niega primera denuncia en contra de Pablo Celi, contralor
9	Asambleístas rinden versión en caso por delincuencia organizada
10	Cómo funciona el Programa de Protección a Víctimas y Testigos en Ecuador

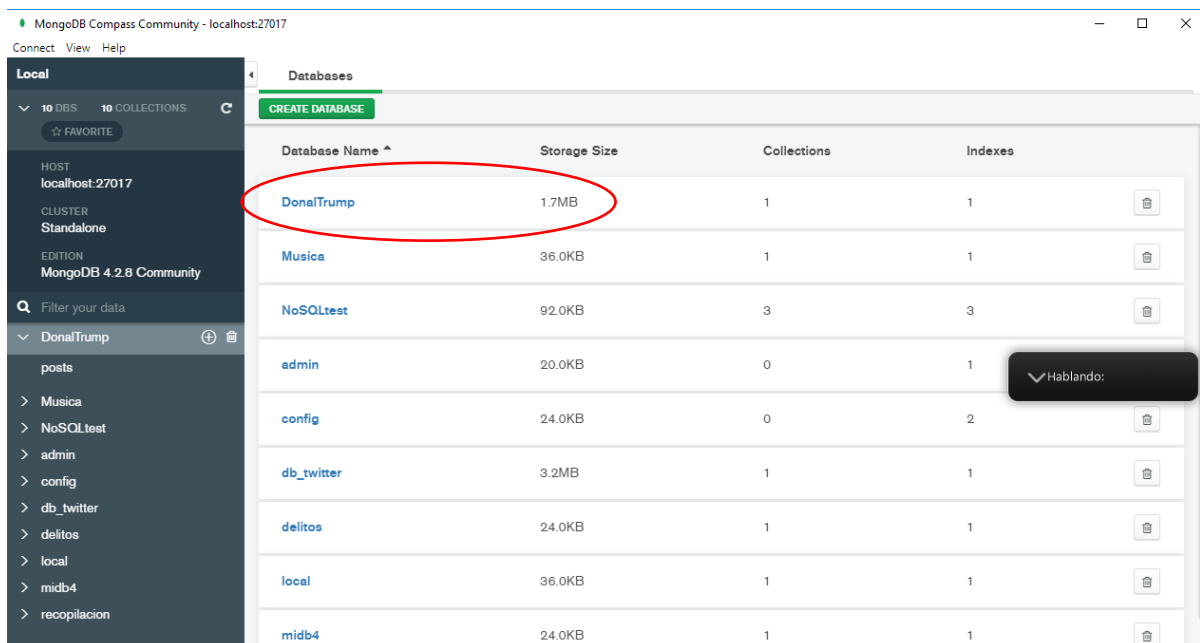
[CANCEL](#) [IMPORT](#)



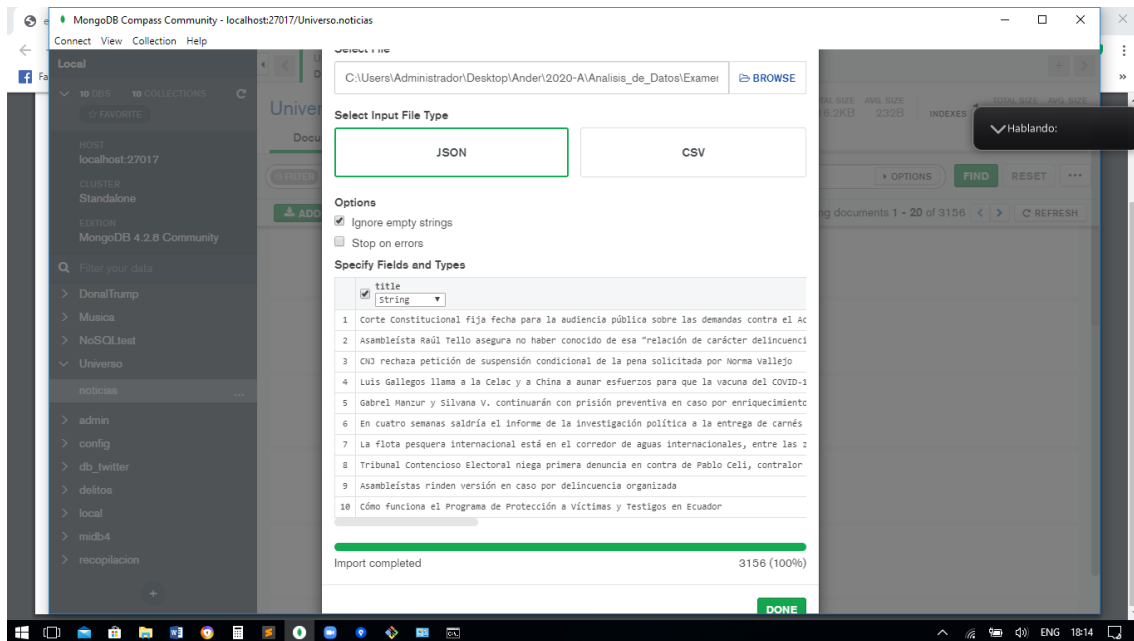
13. Realizamos y ejecutamos el script de tweets\_donaltrump.py con las credenciales de nuestro api



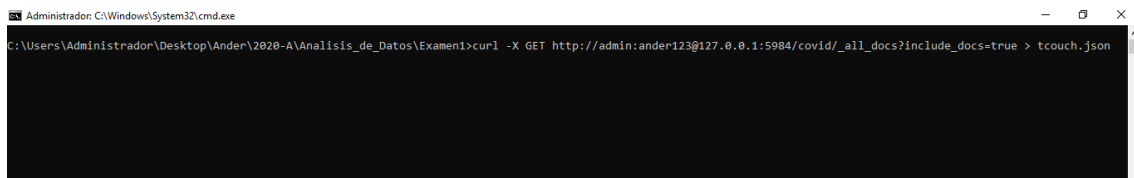
#### 14. Comprobamos en nuestra base de datos de mongo



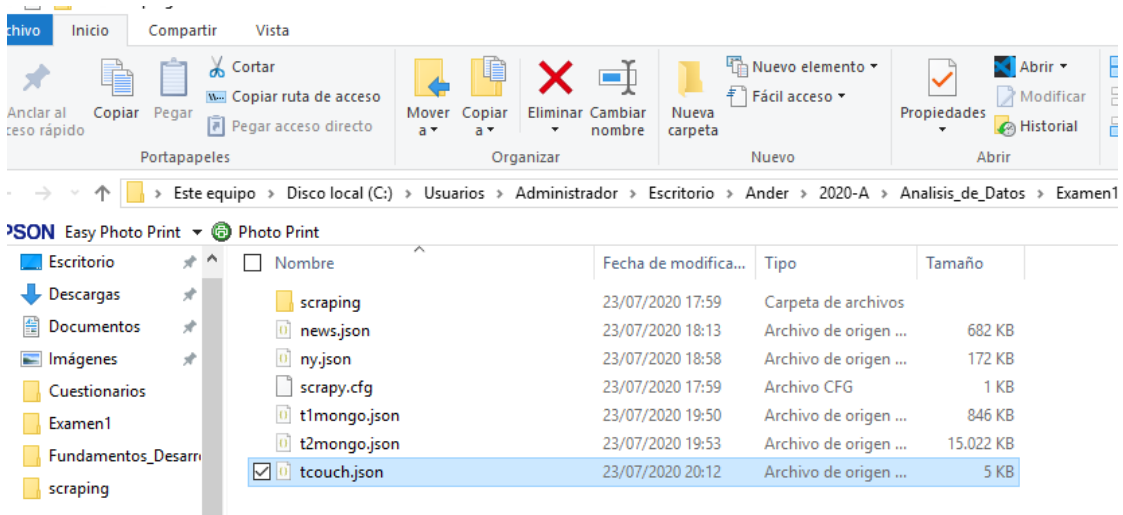
15. Como ya obtuvimos el json de nuestro scraping, lo cargamos a la nueva base de datos DonalTrump para obtener los datos de ambas BD's



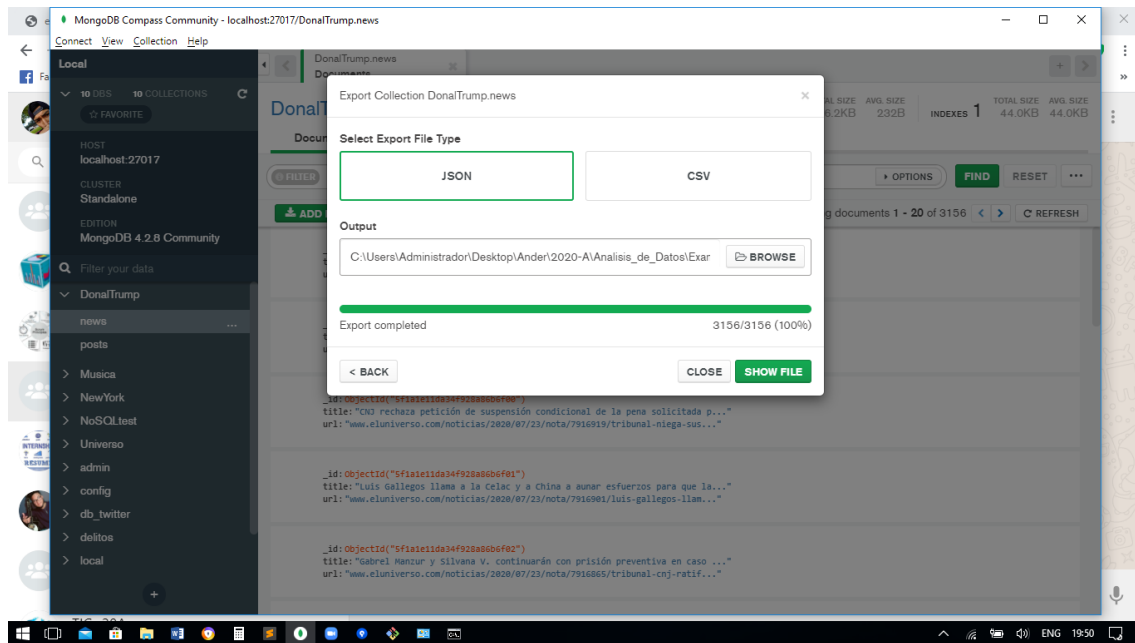
16. Exportamos un json de nuestra base de datos a través del cmd



17. Obteniendo el json tcouch.json

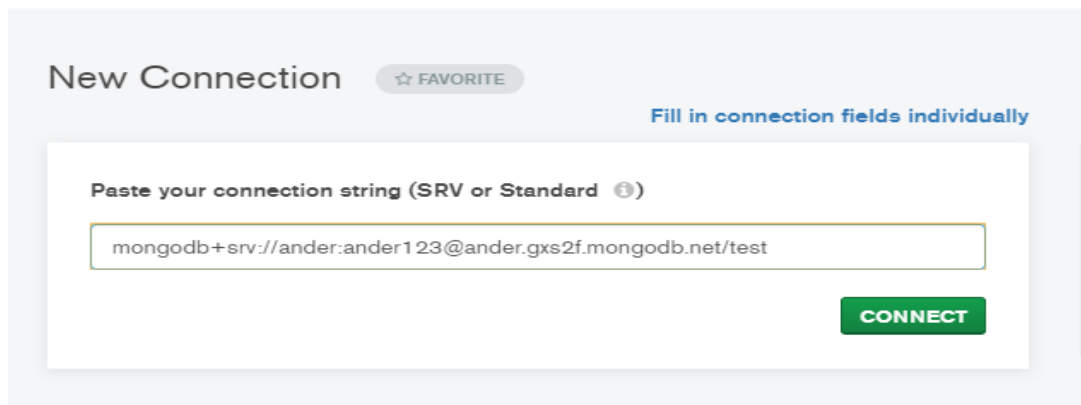


18. Exportamos los datos de nuestra base de mongodb a documentos json



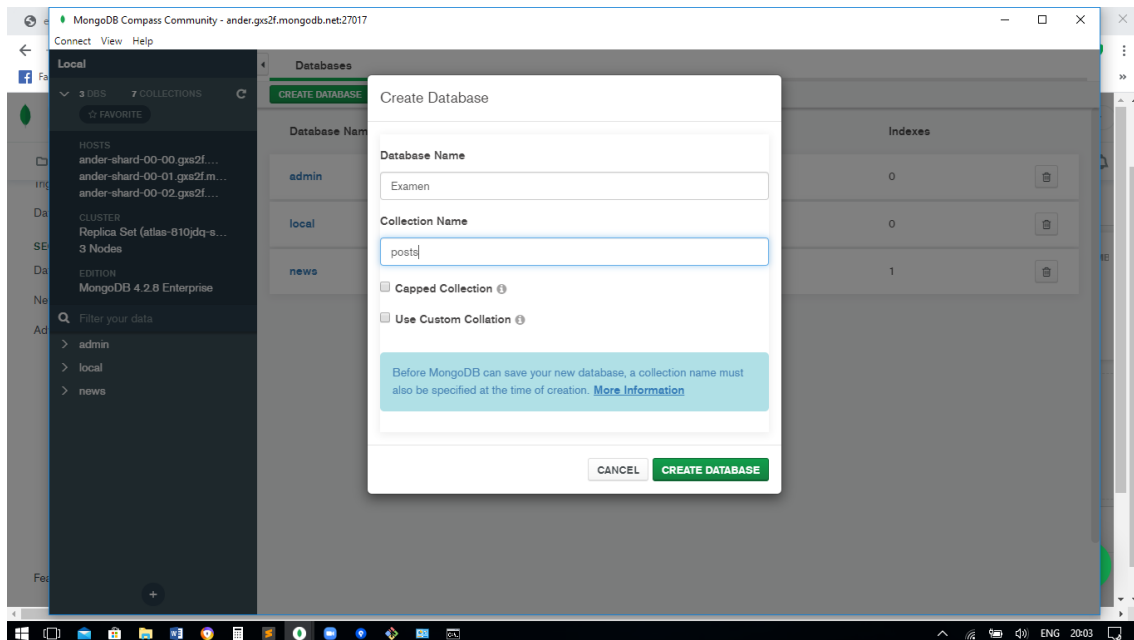
19. Nos conectamos a nuestro cluster en Atlas

:

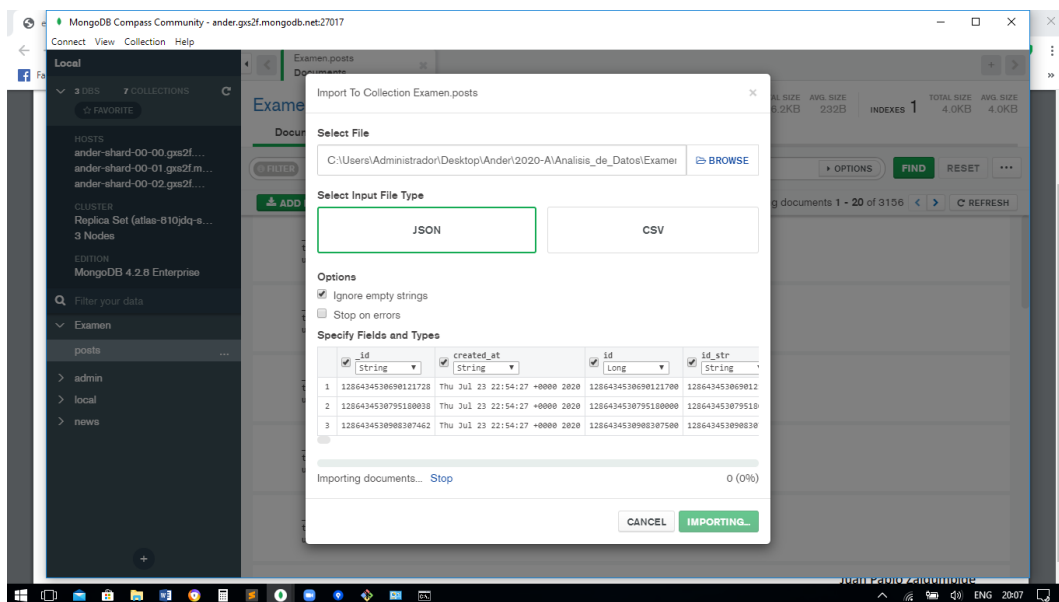


20. Creamos una base de datos en la nube





21. Importamos todos nuestros json's a nuestra base en atlas



22. Enlace de conexión al cluster →  
 mongodb+srv://ander:ander123@ander.gxs2f.mongodb.net/test