***Capturas DataHub***

Se da de alta el nombre del proyecto y el nombre de la base de datos en una colección de MongoDB.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

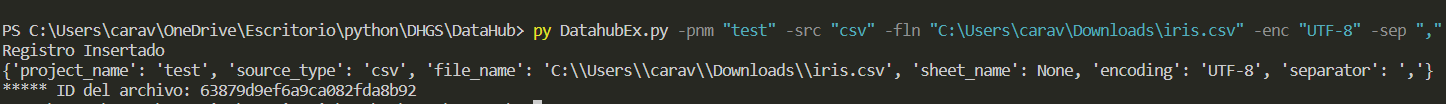
Se crea la base de datos con el mismo nombre que se dio de alta, en ese caso “test”. Igualmente se crean dos colecciones vacías *DataCleaned* y *DataFlows.*

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

A continuación, ejecutamos el comando:

***py DatahubEx.py -pnm "IDS" -src "csv" -fln "C:\Users\path\to\data.csv" -enc "UTF-8" -sep ","***

en la consola para cargar los datos deseados.

Tras haber hecho esto y si la ejecución fue exitosa se nos crearán 3 nuevas colecciones en MongoDB.

Tabla

Descripción generada automáticamente

Si nos vamos a la colección con nombre *DataSource\_Loads* podremos observar que se tiene un registro, esto nos indica que se ha subido con éxito la fuente de datos a nuestra base de datos de MongoDB.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Igualmente, si nos vamos a la colección *DataLoads* podremos observar las filas de nuestra fuente de datos que recién cargamos.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Por último, podemos inspeccionar la colección *DataPerf* donde podremos encontrar un perfilado de las columnas de nuestros datos (medidas de tendencia central, valores máximos y mínimos, tipo de dato, etc.)

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Si queremos aplicar un flujo de datos a nuestra base de datos tendremos que darlo de alta previamente en la colección *DataFlows*.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Para aplicar nuestro flujo creado ejecutamos en la consola el siguiente comando:

***py DatahubTr.py -pnm "IDS" -sid "62ed8d98df7437da17bf6e95" -fid "62ed8e661450a63050065dc8"***

Texto

Descripción generada automáticamente

Si todo ha salido bien, ahora podremos encontrar nuestra base de datos que ha recibido tratamiento en la colección *DataCleaned.*

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Por último, para guardar nuestra base de datos tratada ejecutamos el siguiente comando en consola:

***py DataHubLd.py -pnm "IDS" -sid "62ed8d98df7437da17bf6e95" -otp "csv" -onm "salida - test2"***

Texto

Descripción generada automáticamente

Y podremos encontrar nuestra base de datos tratada con el que le hayamos otorgado, en este caso *test\_processed*.



Cabe destacar que si alguno de los procesos ejecutados por el usuario utilizando el DataHub falla, estos serán cargados a la colección *Log\_Errors.*

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente