

ETL CCE



ALFONSOCAROP@GMAIL.COM



300 2055093

BOGOTA DC, 2018

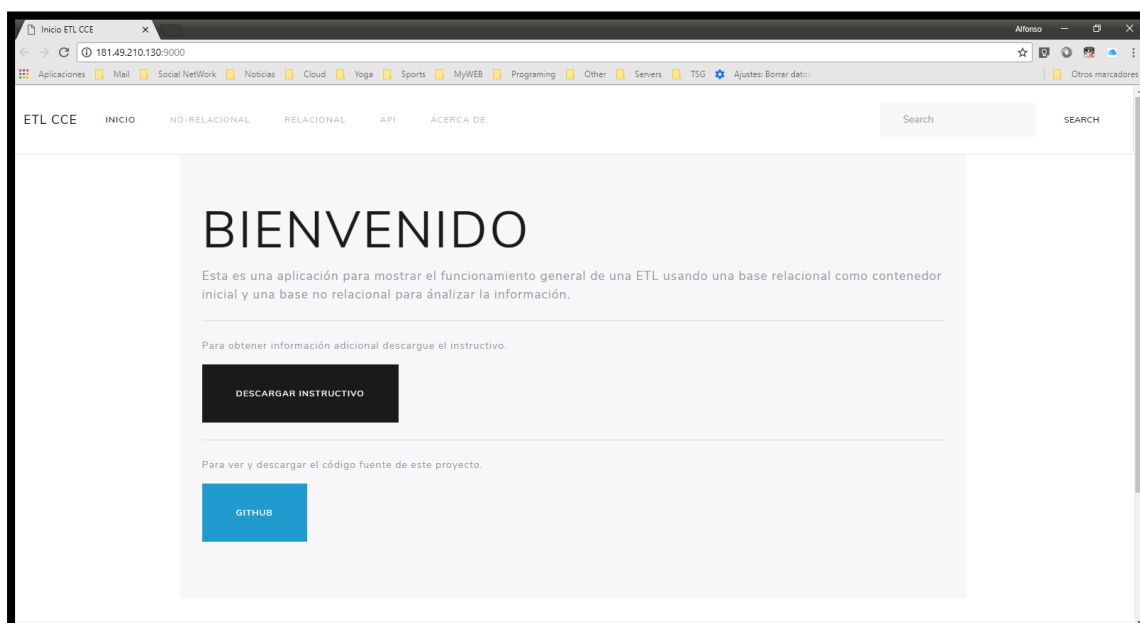
INSTRUCTIVO Y CONTROL DE ARTEFACTOS ALFONSO CARO PORTILLO

Este instructivo tiene como objetivo mostrar el funcionamiento y enumerar los artefactos entregables para la aplicación ejemplo de un ETL y API funcional para Colombia Compra Eficiente usando Python como lenguaje de programación y Django como framework de desarrollo.

I. Ingreso a la aplicación:

Para el ingreso a la aplicación se ingresa desde el navegador la dirección <http://alfonsocarop.com/etl/>

En ella encontraremos el inicio de la aplicación con una breve explicación de su finalidad y dos botones para descargar este instructivo y para ver el código fuente en GitHub respectivamente.

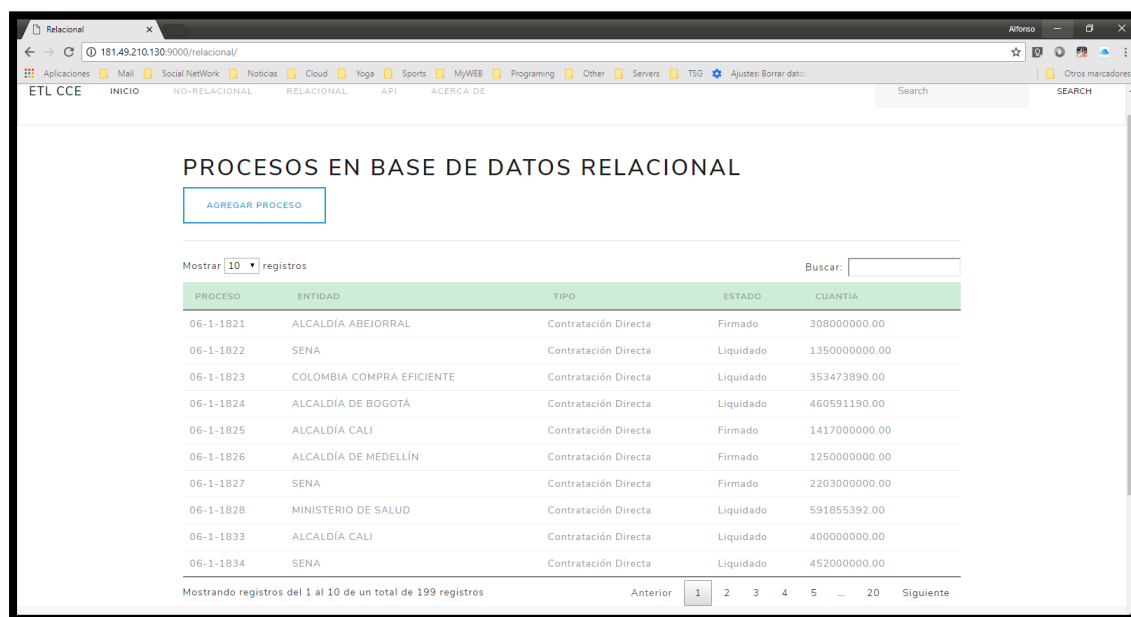


2. ETL.

Para el desarrollo de la ETL se uso como base de datos relacional inicial MySQL, precargada con información referente a Procesos y como concentrador final una base de datos en MongoDB, aunque existen diversas herramientas ETL en el mercado tanto de pago como gratuitas, para este ejemplo se usa código Python nativo para realizar el proceso de extracción, transformación y carga, de esta forma aparte de pedagógica se puede personalizar totalmente el proceso de transporte y transformación de la información.

Paso I. Visualización de la información desde la base de datos relacional en este caso en MySQL.

En el menú principal seleccionamos la opción RELACIONAL.



The screenshot shows a web application with a navigation menu at the top containing 'ETL CCE', 'INICIO', 'NO-RELACIONAL', 'RELACIONAL', 'API', and 'ACERCA DE'. The 'RELACIONAL' option is selected. The main content area is titled 'PROCESOS EN BASE DE DATOS RELACIONAL' and includes a button 'AGREGAR PROCESO'. Below this is a table with 5 columns: 'PROCESO', 'ENTIDAD', 'TIPO', 'ESTADO', and 'CUANTIA'. The table displays 10 records. At the bottom, there is a pagination bar showing 'Mostrando registros del 1 al 10 de un total de 199 registros' and navigation links for 'Anterior', '1', '2', '3', '4', '5', '20', and 'Siguiete'.

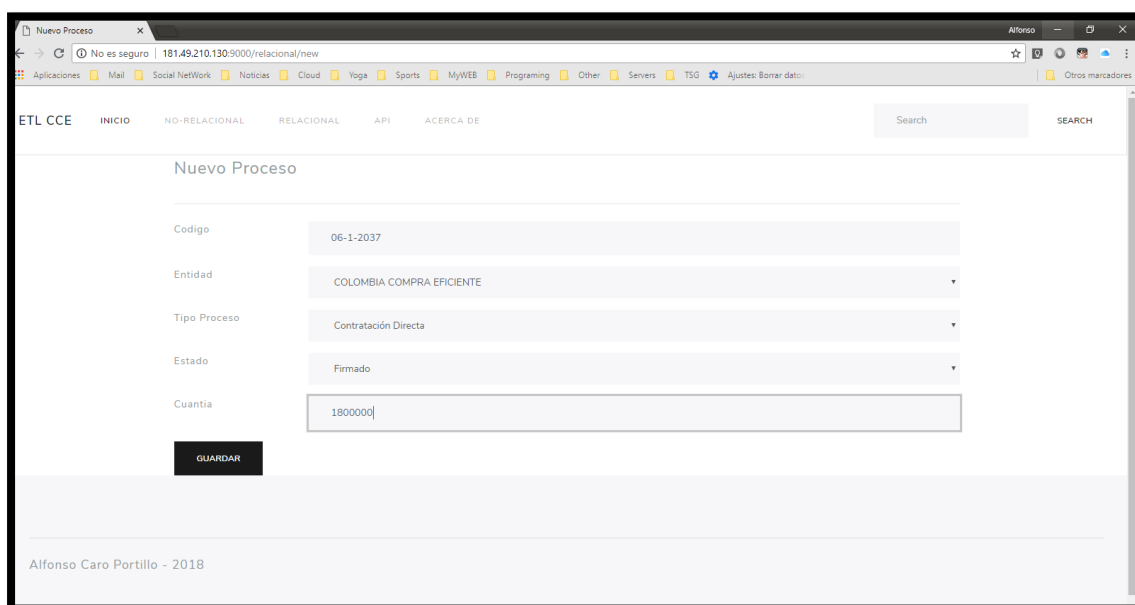
PROCESO	ENTIDAD	TIPO	ESTADO	CUANTIA
06-1-1821	ALCALDÍA ABEJORRAL	Contratación Directa	Firmado	308000000.00
06-1-1822	SENA	Contratación Directa	Liquidado	1350000000.00
06-1-1823	COLOMBIA COMPRA EFICIENTE	Contratación Directa	Liquidado	353473890.00
06-1-1824	ALCALDÍA DE BOGOTÁ	Contratación Directa	Liquidado	460591190.00
06-1-1825	ALCALDÍA CALI	Contratación Directa	Firmado	1417000000.00
06-1-1826	ALCALDÍA DE MEDELLÍN	Contratación Directa	Firmado	1250000000.00
06-1-1827	SENA	Contratación Directa	Firmado	2203000000.00
06-1-1828	MINISTERIO DE SALUD	Contratación Directa	Liquidado	591855392.00
06-1-1833	ALCALDÍA CALI	Contratación Directa	Liquidado	400000000.00
06-1-1834	SENA	Contratación Directa	Liquidado	452000000.00

Paso 2. Agregar Información

Para agregar un nuevo registro a la base de datos damos clic en agregar proceso, esto nos permitirá visualizar un formulario para agregar un nuevo proceso.



A continuación, diligenciamos el formulario con la información de un nuevo proceso y damos clic en el botón guardar.

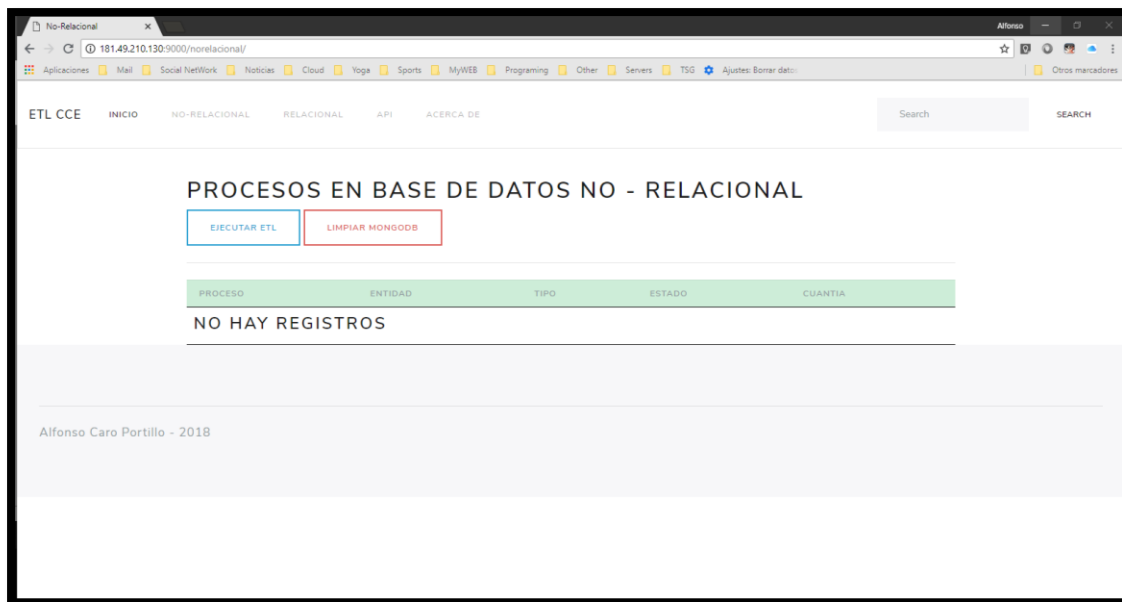
Captura de pantalla de un navegador web mostrando el formulario "Nuevo Proceso". El formulario contiene los siguientes campos: "Codigo" con el valor "06-1-2037", "Entidad" con el valor "COLOMBIA COMPRA EFICIENTE", "Tipo Proceso" con el valor "Contratación Directa", "Estado" con el valor "Firmado", y "Cuantia" con el valor "1800000". Debajo de los campos hay un botón "GUARDAR". En la parte superior del navegador se ven las pestañas "Aplicaciones", "Mail", "Social NetWork", "Noticias", "Cloud", "Yoga", "Sports", "MyWEB", "Programing", "Other", "Servers", "TSG" y "Ajustes: Borrar datos:". En la parte inferior del navegador se ve el texto "Alfonso Caro Portillo - 2018".

De esta forma un nuevo registro añadido a la base de datos relacional y podemos verla en la tabla de datos

Captura de pantalla de una interfaz web que muestra una tabla de datos. La interfaz tiene un título "PROCESOS EN BASE DE DATOS RELACIONAL" y un botón "AGREGAR PROCESO". Debajo del botón hay un campo "Mostrar" con el valor "10" y un campo "Buscar" con el valor "2037". La tabla tiene cinco columnas: "PROCESO", "ENTIDAD", "TIPO", "ESTADO" y "CUANTIA". La tabla muestra un solo registro con los valores: "06-1-2037", "COLOMBIA COMPRA EFICIENTE", "Contratación Directa", "Firmado" y "1800000.00". Debajo de la tabla hay un texto "Mostrando registros del 1 al 1 de un total de 1 registros (filtrado de un total de 201 registros)" y botones "Anterior", "1" y "Siguiente".

Paso 3. Extraer la información de la base de datos

Al dar clic en la pestaña NO RELACIONAL evidenciamos que la base de datos de destino se encuentra vacía tal como la muestra la tabla de visualización.



En esta vista evidenciamos dos botones uno para ejecutar el proceso de extracción y otro para limpiar la base de datos de destino, al dar clic en EJECUTAR ETL se desencadena un proceso on-line que extrae la información de la base de datos Mysql la transforma para que se convierta en formato JSON y se carga a la base de datos de destino MongoDB.



Internamente un proceso en Python completa el ciclo ETL

```
def etl_mongodb(request):
    try:
        # Extract: extraer.
        procesos_sql = Proceso.objects.all()
        for proceso_s in procesos_sql:
            # Transform: transformar.
            entidad_s = proceso_s.CodigoEntidad
            proceso_mc = Proceso_mdb(
                id=str(proceso_s.CodigoProceso),
                codigo=str(proceso_s.CodigoProceso),
                entidad=[{"codigo": str(entidad_s.Codigo), "nombre": str(entidad_s.Nombre)}],
                tipo=str(proceso_s.ID_TipoProceso),
                estado=str(proceso_s.ID_EstadoProceso),
                cuantia=str(proceso_s.Cuantia)
            )
            # Load: cargar.
            try:
                proceso_mc.save()
            except Exception:
                print("error Save")
        except Exception:
            print("error ETL")
    return redirect('norelacional')
```

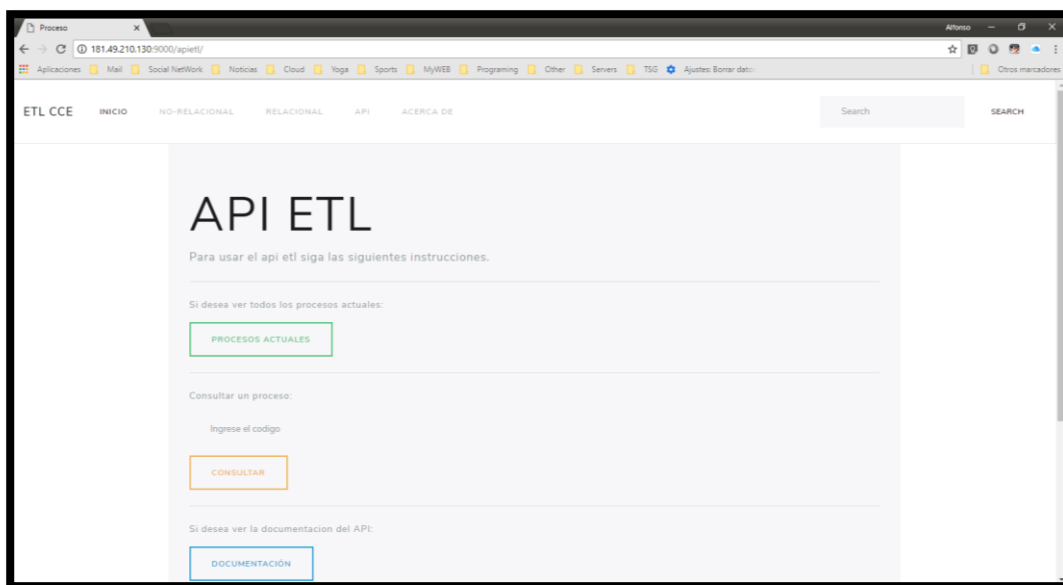
Paso 4. Ver información después de ejecutar ETL.

Al finalizar el proceso de extracción, transformación y carga, se visualiza la información en la tabla con todos los registros iniciales que incluyen los que se han agregado desde la aplicación.

PROCESO	ENTIDAD	TIPO	ESTADO	CUANTIA
06-1-1821	[<Entidad_mdb: ALCALDÍA ABEJORRAL>]	Contratación Directa	Firmado	308000000.00
06-1-1822	[<Entidad_mdb: SENA>]	Contratación Directa	Liquidado	1350000000.00
06-1-1823	[<Entidad_mdb: COLOMBIA COMPRA EFICIENTE>]	Contratación Directa	Liquidado	353473890.00
06-1-1824	[<Entidad_mdb: ALCALDÍA DE BOGOTÁ>]	Contratación Directa	Liquidado	460591190.00
06-1-1825	[<Entidad_mdb: ALCALDÍA CALI>]	Contratación Directa	Firmado	1417000000.00
06-1-1826	[<Entidad_mdb: ALCALDÍA DE MEDELLÍN>]	Contratación Directa	Firmado	1250000000.00
06-1-1827	[<Entidad_mdb: SENA>]	Contratación Directa	Firmado	2203000000.00
06-1-1828	[<Entidad_mdb: MINISTERIO DE SALUD>]	Contratación Directa	Liquidado	591855392.00
06-1-1833	[<Entidad_mdb: ALCALDÍA CALI>]	Contratación Directa	Liquidado	400000000.00
06-1-1834	[<Entidad_mdb: SENA>]	Contratación Directa	Liquidado	452000000.00

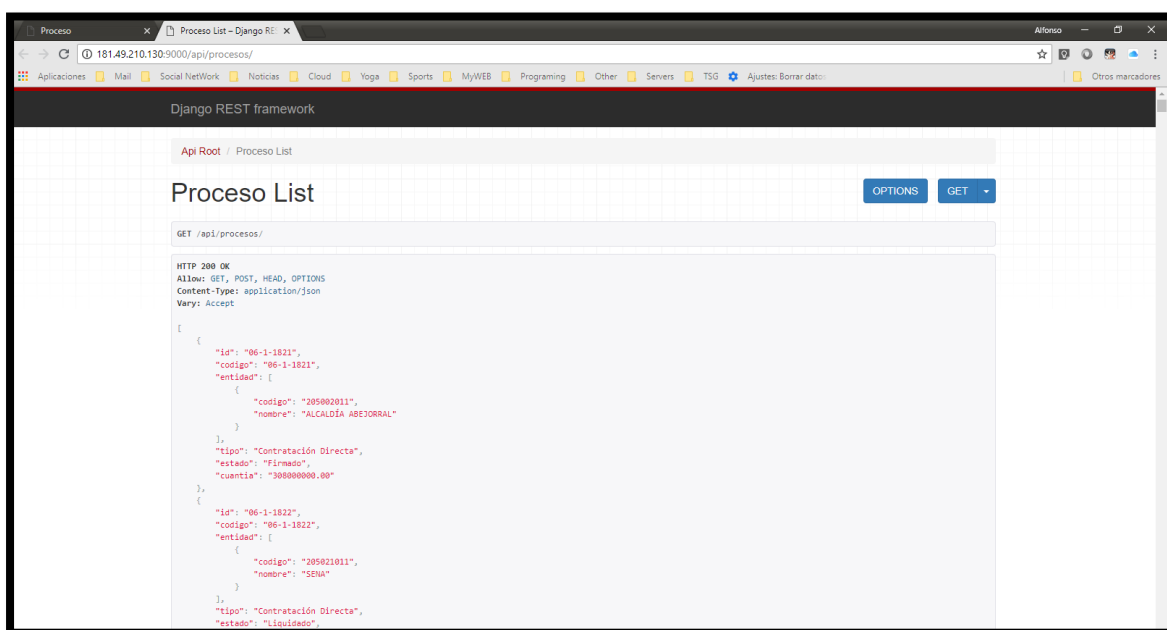
3. API.

Como funcionalidad adicional la aplicación cuenta con un api rest que permite la consulta, creación, actualización y eliminación de registros desde peticiones HTTP a la base de datos no relacional. En la pestaña API de la aplicación se visualizan las siguientes opciones:

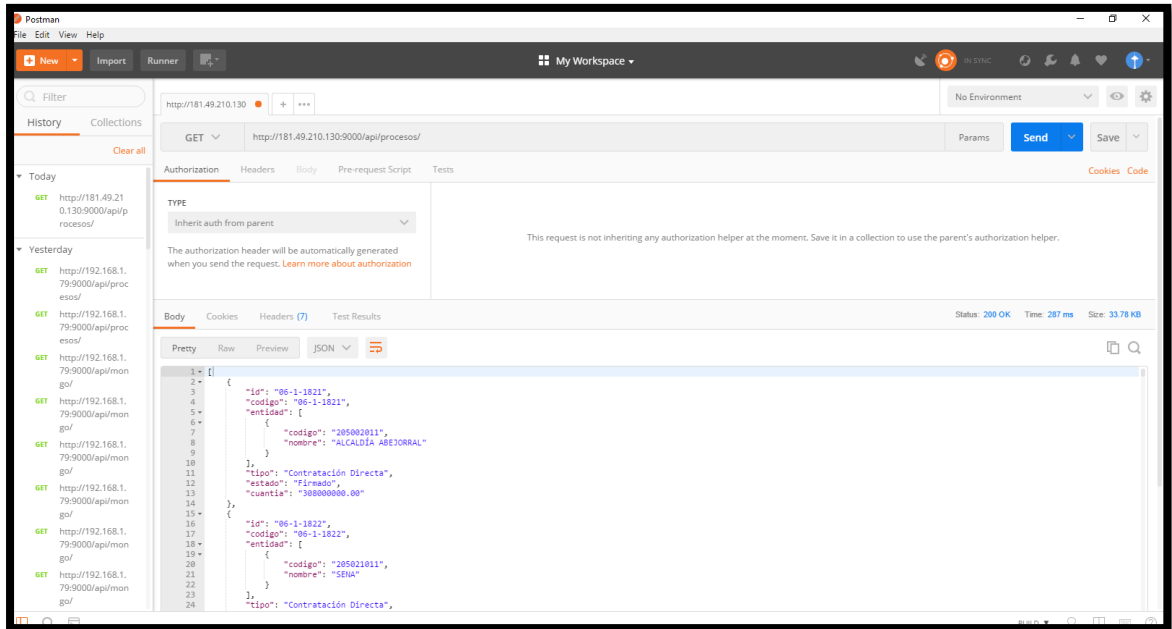


a. Listar registros.

Al dar clic en el botón Procesos actuales la aplicación nos redirigirá a la pagina del API donde se listaran todos los procesos que fueron migrados desde la base de datos Relacional a MongoDB.

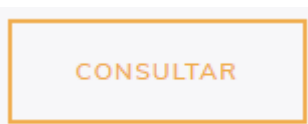


Usando un cliente http de API REST como POSTMAN podemos ver el funcionamiento del api y su resultado exitoso en formato JSON.



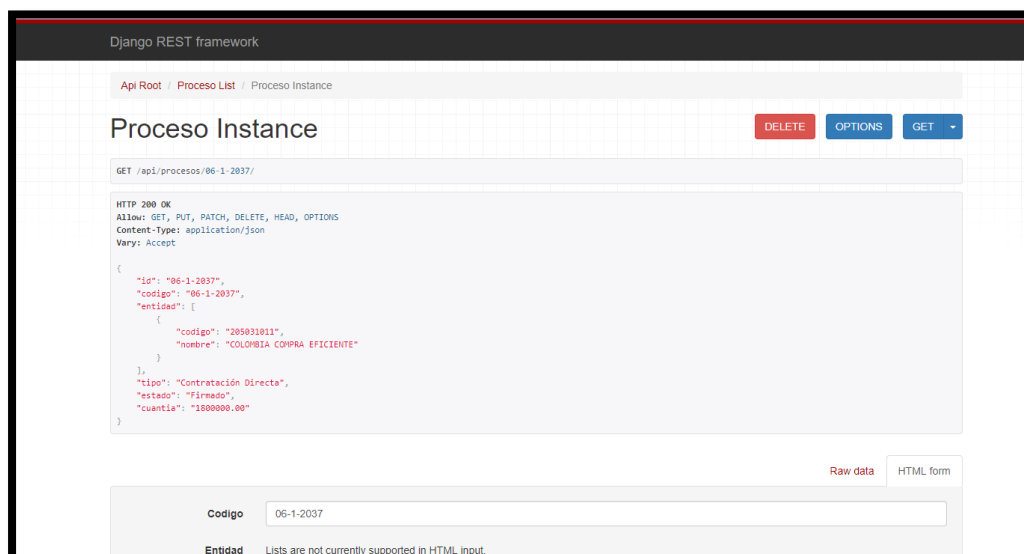
b. Consulta API de procesos.

Para realizar la consulta de un proceso en específico usando el API, la aplicación provee una caja de texto y botón que redirecciona al usuario al resultado esperado.



Consultar un proceso:

CONSULTAR



c. Documentación.

DOCUMENTACIÓN

Para visualizar la documentación del API y su funcionamiento detallado damos clic en el botón DOCUMENTACION y la aplicación nos redireccionara a una pagina con toda la información correspondiente.

The screenshot displays the ETL CCE API documentation interface. On the left, a sidebar shows the 'ETL CCE' logo and navigation links for 'proceso' and 'procesos'. The main content area is titled 'ETL CCE' and 'proceso'. It features two sections: 'read' and 'create'. Each section includes a method (GET for read, POST for create), a path, and a table of path parameters. The 'read' section shows a parameter 'id' with a 'required' status. The 'create' section shows a parameter 'id' with a 'required' status. Below the 'create' section, there is a 'Request Body' section indicating that the request body should be a 'application/json' encoded object. To the right of each section, there are terminal snippets showing how to use the 'coreapi' CLI to interact with the API.

Parameter	Description
id required	

Parameter	Description
id required	

2018

ALFONSO CARO PORTILLO