

CODERFLEX

# Data Engineering

## Consigna de Segundo Pre entregable

***CODERHOUSE***

# Carga de datos en Amazon Redshift

Para finalizar tu segundo pre entregable, te proponemos que el script de la entrega 1 adapté los datos leídos de la API y cargarlos en la tabla creada en la pre-entrega anterior en Redshift.

## Objetivos

### Objetivos generales

- El script de la entrega 1 deberá adaptar datos leídos de la API y cargarlos en la tabla creada en la pre-entrega anterior en Redshift de forma eficiente. En esta entrega se hace énfasis en la limpieza de los datos crudos obtenidos de la API

### Objetivos específicos

- Generar ETLs a partir de información de APIs usando las librerías requests, json, psycopg2/SqlAlchemy y pandas
- Solucionar una situación real de ETL donde puedan llegar a aparecer duplicados, nulos y valores atípicos durante la ingesta-Transformación- Carga de la data

## Requisitos

Este trabajo cuenta con dos instancias que se deben **mostrar dentro de una misma presentación**:

1

[Código/Script en Python](#)

2

[Tabla en Amazon Redshift](#)

## 1. Código/Script

**Código en Python subido ya sea en repositorio de GitHub o en Google Drive. Tabla creada en Redshift con los datos de muestra que hayan sido cargados mediante el script**

El objetivo es tener un código inicial que será usado en el proyecto final como un MVP de ETL . El script debería extraer datos en JSON desde una API pública o gratuita para luego convertir estos datos en diccionario Python y posterior manipulación. Considera los siguientes puntos:

- **El código debería utilizar sólo un dataframe** Es decir, no se debería estar utilizando más de un solo dataframe para subir los datos

Recuerda que esto será la base para tu proyecto final.

## 2. Tabla en Amazon Redshift

**Tabla creada en Redshift con los datos de muestra que hayan sido cargados mediante el script.**

A su vez, la entrega involucra la creación de una versión inicial de la tabla donde los datos serán cargados posteriormente. Considera los siguientes puntos para su elaboración:

1. **Los datos deben ser extraídos y cargados con sus correspondientes tipos de datos en relación a la tabla creada en Redshift.**
2. **Todas las columnas deberían ser cargadas en la tabla.**
3. **Debe haber una clave primaria compuesta definida en la tabla o en el código**
4. **En caso de que se quiera insertar una fila con los mismos datos, debe ser reemplazada por los nuevos? Por ejemplo: la columna "fecha" y "ciudad" puede ser una clave primaria compuesta, ya que no deberían haber 2 datos diferentes para una misma ciudad en un mismo día.**

Recuerda que esto será la base para tu proyecto final.

# Carga de datos en Amazon Redshift

## Recomendaciones

- Consultar la documentación oficial de Pandas; profundizar en su uso.
- Revisar el instrumento de evaluación

## Ejemplos

Para guiarte, te compartimos el siguiente ejemplo:

- [Ejemplo](#)

## Criterios de evaluación

Para la evaluación de tu Proyecto Final, tendremos en cuenta los siguientes [criterios de evaluación](#).