




# Manual de instalación y ejecución

Sebastian Ramiro Guerrero Fierro, Nestor  
Martin Lozada Cueva y Jose Ignacio Escudero  
Rodriguez



## Tabla de Contenidos

<i>¿Cómo instalar y ejecutar exitosamente el proyecto? .....</i>	<b>2</b>
<i>1. Requisitos previos .....</i>	<b>2</b>
<i>2. ¿Cómo instalar los paquetes? .....</i>	<b>2</b>
<i>3. ¿Cómo configurar la base de datos? .....</i>	<b>2</b>
<i>4. Scripts SQL.....</i>	<b>3</b>
<i>5. ¿Cómo generamos los datos?.....</i>	<b>3</b>
<i>3. ¿Cómo ejecutar el ETL?.....</i>	<b>3</b>
<i>4. ¿Cómo ejecutar el proyecto integrador?.....</i>	<b>3</b>
<i>Diseño del proyecto .....</i>	<b>4</b>
<i>1. Diagrama área de staging .....</i>	<b>4</b>
<i>2. Diagrama área de SOR.....</i>	<b>5</b>

# ¿Cómo instalar y ejecutar exitosamente el proyecto?

## 1. Requisitos previos

- i. Para esta etapa, se deberían tener instaladas las últimas versiones de las siguientes herramientas:
  - a. Python 3 (Stable release: 3.11.1)
  - b. MySQL con Workbench (Stable release: 8.0.32)
  - c. Power BI Desktop (December 2022 Update (2.112. 283.0))
- ii. Clonación del repositorio con CLI en la terminal.
  - a. Copia y pega el comando: “git clone [https://github.com/xjose97x/proyecto\\_integrador\\_udla\\_bdd3.git](https://github.com/xjose97x/proyecto_integrador_udla_bdd3.git)”
- iii. Abrir el proyecto en IDE de preferencia, se recomienda VS Code.

## 2. ¿Cómo instalar los paquetes?

Para la correcta ejecución de este proyecto Python necesita tener instalados un grupo de paquetes específicos, se recomienda instalar estos paquetes en un entorno virtual.

- i. Librerías pertinentes:
  - a. autopep8
  - b. Faker
  - c. numpy
  - d. pandas
  - e. pycodestyle
  - f. PyMySQL
  - g. python-dateutil
  - h. python-dotenv
  - i. pytz
  - j. si
  - k. SQLAlchemy
  - l. Toml

Para instalarlos ejecutar el siguiente comando en la terminal:

- pip install -r requirements.txt

## 3. ¿Cómo configurar la base de datos?

Para que la base de datos pueda conectarse con el entorno de Python se deberá modificar los archivos de configuración (*core.properties* y *staging.properties*) que se encuentran en el directorio *config* donde se encuentran las configuraciones de la base de datos.

## 4. Scripts SQL

Este grupo de scripts se encuentran en el directorio llamado *sql*.

Es importante recalcar que se deberán ejecutar en un orden determinado.

El orden de ejecución es el siguiente:

1. *sql/init.sql* (se inicializan los esquemas de la base de datos)
2. *sql/dwh\_stg.sql* (se crean las tablas de extracción)
3. *sql/dwh\_tra.sql* (se crean las tablas de transformación)
4. *sql/dwh.sql* (se crean las tablas para el ambiente SOR)

## 5. ¿Cómo generamos los datos?

1. Datos obtenidos de Pronaca:
  - *Compra\_productos.csv* (contiene un desglose de las compras por mes de distintos productos)
  - *Compras\_consolidadas\_2021\_2022.csv* (contiene diferencia de compras entre el año 2021 y 2022 de distintas categorías de productos)
  - *Riesgo\_proveedores\_2021.csv* (contiene el riesgo que conlleva cada proveedor del año 2021)
2. Datos generados con Faker:
  - *Gatos.csv* (contiene todos los gatos realizados por la empresa en los años 2021 y 2022)
  - *Ingresos.csv* (contiene todos los ingresos realizados por la empresa en los años 2021 y 2022)
  - *Riesgo\_proveedores.csv* (se alimenta de riesgos proveedores 2021 y lo extiende generando datos del 2022)

## 3. ¿Cómo ejecutar el ETL?

Ejecutar en la terminal en siguiente comando:

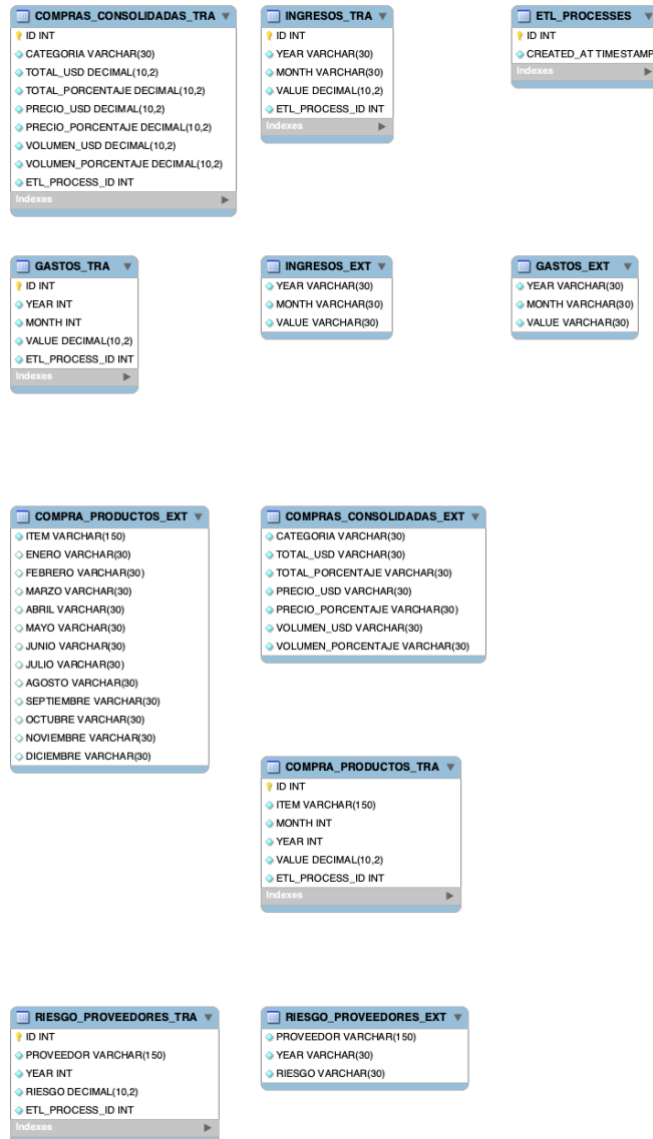
- `python py_startup.py`

## 4. ¿Cómo ejecutar el proyecto integrador?

Abrir el documento *powerbi/POWER\_BI..Analisis.pbix* con Power BI Desktop.

# Diseño del proyecto

## 1. Diagrama área de staging



## 2. Diagrama área de SOR

COMPRA_PRODUCTOS
ID INT
ITEM VARCHAR(150)
MONTH INT
YEAR INT
VALUE DECIMAL(10,2)
Indexes

GASTOS
ID INT
YEAR INT
MONTH INT
VALUE DECIMAL(10,2)
Indexes

RIESGO_PROVEEDORES
ID INT
PROVEEDOR VARCHAR(150)
YEAR INT
RIESGO DECIMAL(10,2)
Indexes

INGRESOS
ID INT
YEAR VARCHAR(30)
MONTH VARCHAR(30)
VALUE DECIMAL(10,2)
Indexes

COMPRAS_CONSOLIDADAS
ID INT
CATEGORIA VARCHAR(30)
TOTAL_USD DECIMAL(10,2)
TOTAL_PORCENTAJE DECIMAL(10,2)
PRECIO_USD DECIMAL(10,2)
PRECIO_PORCENTAJE DECIMAL(10,2)
VOLUMEN_USD DECIMAL(10,2)
VOLUMEN_PORCENTAJE DECIMAL(10,2)
Indexes