



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

**MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR  
OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE  
SAN CARLOS DE GUATEMALA**

**Jefferson José Carlos Linares Cerón**  
Asesorado por Ingeniero Herman Igor Veliz Linares

Guatemala, septiembre de 2022



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR  
OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE  
SAN CARLOS DE GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**JEFFERSON JOSÉ CARLOS LINARES CERÓN**  
ASESORADO POR INGENIERO HERMAN IGOR VELIZ LINARES

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, SEPTIEMBRE DE 2022



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martinez
VOCAL III	Ing. José Milton De León Bran
VOCAL IV	Br. Kevin Vladimir Armando Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Ing. Sergio Leonel Gómez Bravo
EXAMINADOR	Ing. Carlos Alfredo Azurdia Morales
EXAMINADORA	Inga. Floriza Felipa Avila Pesquera de Medinilla
SECRETARIO	Ing. Hugo Humberto Rivera Pérez



## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR  
OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE  
SAN CARLOS DE GUATEMALA**

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 5 de agosto de 2021.



**Jefferson José Carlos Linares Cerón**



Guatemala, 05 de julio de 2022

Ingeniero Oscar Argueta Hernández  
Director de la Unidad de EPS  
Facultad de Ingeniería, USAC  
Ciudad Universitaria, Guatemala

Estimado Ingeniero Argueta

Por este medio atentamente hago de su conocimiento que como Asesor de la Práctica del Ejercicio Profesional Supervisado (EPS), que el estudiante **Jefferson José Carlos Linares Cerón**, de la carrera de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, quien se identifica con CUI: **3373833162002** y Registro Académico: **201504448**, he revisado el informe final titulado: **“MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA”**.

En tal virtud, **lo doy por aprobado**, solicitándole darle el trámite respectivo, Sin otro particular me es grato suscribirme.

Atentamente



Ing. Herman Igor Veliz Linares  
COLEGIADO No. 4836

Ing. Herman Igor Veliz Linares  
Asesor del proyecto EPS  
Escuela de Ingeniería en Ciencias y Sistemas



Universidad de San Carlos de  
Guatemala



Facultad de Ingeniería  
Unidad de EPS

Guatemala, 22 de julio de 2022.  
REF.EPS.D.228.07.2022.

Ing. Carlos Gustavo Alonzo  
Director Escuela de Ingeniería Ciencias y Sistemas  
Facultad de Ingeniería  
Presente

Estimado Ingeniero Alonzo:

Por este medio atentamente le envío el informe final correspondiente a la práctica del Ejercicio Profesional Supervisado, (E.P.S) titulado **MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, que fue desarrollado por el estudiante universitario **Jefferson José Carlos Linares Cerón, Registro Académico 201504448 y CUI 3373 83316 2002** quien fue debidamente asesorado por el Ing. Herman Igor Veliz Linares y supervisado por la Inga. Floriza Felipa Ávila Pesquera de Medinilla.

Por lo que habiendo cumplido con los objetivos y requisitos de ley del referido trabajo y existiendo la aprobación del mismo por parte del Asesor y la Supervisora de EPS, en mi calidad de Director apruebo su contenido solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,

“Id y Enseñad a Todos”



Ing. Oscar Argueta Hernández  
Director Unidad de EPS

/ra





Universidad San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala 29 de julio de 2022

Ingeniero  
**Carlos Gustavo Alonzo**  
Director de la Escuela de Ingeniería  
En Ciencias y Sistemas

Respetable Ingeniero Alonzo:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación-EPS del estudiante **JEFFERSON JOSÉ CARLOS LINARES CERÓN** carné **201504448** y CUI **3373 83316 2002**, titulado: **“MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA”** y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,



**Ing. Carlos Alfredo Azurdia**  
Coordinador de Privados  
y Revisión de Trabajos de Graduación



UNIVERSIDAD DE SAN CARLOS  
DE GUATEMALA



FACULTAD DE INGENIERÍA

LNG.DIRECTOR.177.EICCSS.2022

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador de área y la aprobación del área de lingüística del trabajo de graduación titulado: **MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por: **Jefferson José Carlos Linares Cerón**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería.

“ID Y ENSEÑAD A TODOS”



Msc. Ing. Carlos Gustavo Alonzo  
Director

Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, septiembre de 2022





Decanato  
Facultad de Ingeniería  
24189101- 24189102  
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.628.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **MIGRACIÓN DE REGISTROS ESTUDIANTILES DEL CENTRO UNIVERSITARIO SUR OCCIDENTE AL NUEVO SISTEMA DE CONTROL ACADÉMICO DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por: **Jefferson José Carlos Linares Cerón**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Inga. Aurelia Anabela Cordova Estrada



Decana

Guatemala, septiembre de 2022

AACE/gaoc



## **ACTO QUE DEDICO A:**

- Dios** Por brindarnos la vida y salud para continuar mejorando día con día.
- Mis padres** Edna Cerón y José Linares, quienes me han apoyado incondicionalmente toda mi vida, trabajando sin descanso para darnos la oportunidad de prepararnos a mí y a mis hermanos, hasta lograr desarrollarnos profesional, moral y éticamente.
- Mis hermanos** Catherine, Brandon, Rachell y Sharon Linares.
- Mis abuelas** Augusta Cerón y Melida Osorio, quienes con amor siempre reciben a sus nietos, y en sus oraciones siempre ponen a la familia.
- Mis amigos** De infancia quienes, a pesar de los años, seguimos compartiendo experiencias, y también a los grandes amigos que esta carrera me permitió conocer y juntos apoyarnos para avanzar en la misma.



## **AGRADECIMIENTOS A:**

<b>Universidad de San Carlos de Guatemala</b>	La casa de estudios de muchos guatemaltecos con sueños por cumplir, y que en ella encuentran una gran oportunidad para conseguirlos.
<b>Familia</b>	Por todo el apoyo brindado, en especial a mis padres.
<b>Mis tías</b>	Amarilis, Blanca, Milvia y Emelina Cerón; Andrea, Flor, Vilma y Miriam Linares, por siempre dar ánimos y consejos para continuar preparándome.
<b>Asesores</b>	Ing. Herman Igor Veliz Linares, Lic. Rodrigo Mendizábal Burastero y Estuardo Tello, por el acompañamiento, guía y revisiones brindadas en todo el proyecto de migración.
<b>Catedráticos</b>	Por todo el conocimiento y experiencia compartida, en el que cada minuto cuenta para la formación de buenos profesionales en el país.



## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	V
LISTA DE SÍMBOLOS .....	VII
GLOSARIO .....	IX
RESUMEN.....	XV
OBJETIVOS.....	XVII
INTRODUCCIÓN .....	XIX
1. FASE DE INVESTIGACIÓN – CUNSUROC .....	1
1.1. Antecedentes de la institución .....	1
1.1.1. Reseña histórica .....	1
1.1.2. Misión .....	1
1.1.3. Visión.....	2
1.1.4. Servicios que realiza.....	2
1.2. Descripción de las necesidades .....	2
1.3. Priorización de las necesidades .....	3
2. FASE TÉCNICO PROFESIONAL .....	5
2.1. Descripción del proyecto .....	5
2.2. Sistemas de bases de datos.....	6
2.2.1. Entidades y vínculos.....	7
2.2.1.1. Entidad .....	7
2.2.1.2. Vínculo.....	7
2.2.2. Propiedades.....	8
2.2.3. Datos y modelos de datos .....	8
2.2.3.1. Datos .....	8

2.2.3.2.	Modelo de datos.....	8
2.2.4.	SQL.....	9
2.2.4.1.	Tipos de datos.....	9
2.2.4.1.1.	DDL.....	9
2.2.4.2.	DML.....	11
2.2.5.	Migración de bases de datos.....	12
2.2.5.1.	Categorías.....	12
2.2.5.1.1.	Migración de base de datos .....	12
2.2.5.1.2.	Migración de aplicación.....	12
2.2.5.1.3.	Migración de almacenamiento.....	13
2.2.6.	FoxBase Proffesional versión 2.5.....	13
2.2.7.	MySQL .....	14
2.2.8.	ETL.....	14
2.3.	Solución del proyecto .....	15
2.3.1.	Base de datos destino.....	18
2.3.2.	Información migrada.....	22
2.3.3.	Sistema de control académico WEB .....	24
2.4.	Costos del proyecto.....	26
2.5.	Beneficios del proyecto .....	27
3.	FASE DE CAPACITACIÓN.....	31
3.1.	Capacitación propuesta del Sistema de Control Académico.....	31
3.1.1.	Manual de usuario.....	31
3.2.	Capacitación algoritmo ETL .....	33
3.2.1.	Manual técnico y documentación asociada.....	35
	CONCLUSIONES.....	37
	RECOMENDACIONES .....	39

BIBLIOGRAFÍA.....	41
APÉNDICES .....	43
ANEXO .....	57



## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1.	Diagrama de Modelo Relacional de la base de datos destino.....	19
2.	Contenedores de base de datos local.....	20
3.	Diagrama de flujo de proceso ETL general.....	20
4.	Diagrama de flujo del proceso de migración.....	21
5.	Proporción de registros migrados.....	23
6.	Consulta de cursos aprobados por estudiante y carrera.....	24
7.	Consulta de cursos asignados por estudiante.....	25
8.	Vista de cursos aprobados por el estudiante y carrera.....	26
9.	Repositorio privado del CUNSUROC.....	34

### TABLAS

I.	Estructura de un acta.....	16
II.	Resumen de registros migrados.....	22
III.	Costos asociados al proyecto.....	27



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
<b>*.sql</b>	Archivos con instrucciones del lenguaje estructurado de consulta.
<b>*.json</b>	Archivos del formato JSON el cual es representado por una estructura basada en la definición de objetos en Javascript.
<b>*.dbf</b>	Archivos que contienen información de registros de una base de datos proveniente de FoxBase.
<b>YYYY-MM-DD</b>	Formato de fecha en SQL
<b>hh:mm:ss</b>	Formato de hora en SQL
<b>[,name_fieldN]*</b>	Hace referencia a una expresión regular que indica la multiplicidad de 0 a N nombres pertenecientes a SQL.
<b>[,valueN]*</b>	Hace referencia a una expresión regular que indica la multiplicidad de 0 a N valores pertenecientes a SQL.
<b>&gt;</b>	Operador relacional mayor que
<b>;</b>	Punto y coma; significa el final de una instrucción SQL



## GLOSARIO

<b><i>ADD</i></b>	Instrucción SQL.
<b><i>ALTER</i></b>	Instrucción SQL.
<b><i>ANSI/ISO SQL</i></b>	Del inglés (American National Standards Institute / International Organization of Standardization), Instituto Nacional Americano de Estándares y la Organización Internacional de Estandarización, son instituciones y organizaciones que se encargan de establecer estándares en este caso relacionadas al lenguaje estructurado de consulta SQL.
<b><i>Byte</i></b>	Unidad de almacenamiento informático equivalente a 8 bits.
<b><i>CMD</i></b>	Símbolo de sistema de Windows.
<b><i>CREATE</i></b>	Instrucción SQL.
<b><i>CSV</i></b>	Del inglés (Comma-Separated Values), son archivos los cuales se conforma por n cantidad de líneas que representan registros y estos registros a su vez se encuentran separados por comas.
<b><i>CUNSUROC</i></b>	Centro Universitario Sur Occidente.

<b><i>DATE</i></b>	Tipo de dato SQL.
<b><i>DBMS</i></b>	Del inglés (Data Base Management System), sistema administrador de bases de datos, permite la gestión (creación, modificación y eliminación), de diferentes objetos dentro de las bases de datos.
<b><i>DDL</i></b>	Del inglés (Data Definition Language), es el lenguaje de definición de datos, ofrece diferentes sentencias para creación de diferentes objetos dentro de la base de datos.
<b><i>DECIMAL</i></b>	Tipo de dato SQL.
<b><i>DELETE</i></b>	Instrucción SQL.
<b><i>DML</i></b>	Del inglés (Data Manipulation Language), es el lenguaje de manipulación de datos, que permite crear, consultar, modificar, eliminar la estructura de los registros almacenados en los objetos de las bases de datos.
<b><i>Docker</i></b>	Es una plataforma de código abierto que permite desarrollar, y ejecutar aplicaciones de software, permite la entrega de software de una forma mucho más rápida, ya que permite manejar la infraestructura necesaria para cada aplicación, teniendo como base sus contenedores: que representan imágenes de

sistemas operativos como Linux, en la que se despliegan dichas aplicaciones.

***DROP***

Instrucción SQL.

***ETL***

Del inglés (Extract, Transform and Load), proceso para ejecución de migraciones de bases de datos.

***Excel***

Microsoft Excel es una hoja de cálculo propiedad de Microsoft, utilizado para creación de cálculos estadísticos, representación de gráficos, tablas, entre otros.

***FoxBase***

Lenguaje de programación orientado a procedimientos y también un sistema gestor de bases de datos.

***FROM***

Instrucción SQL.

***Gb***

*Giga Bite* unidad de almacenamiento informático equivalente a 1 000 Mb.

***Git***

Sistema de control de versiones que permiten unificar archivos generalmente con código en cualquier lenguaje de programación, con el fin de actualizar versiones de un software en específico.

<b>GITHUB</b>	Servicio en la nube que permite gestionar un sistema de control de versiones como git para manejo de repositorios en el que se puede desarrollar software de una manera más ordenada y eficaz.
<b>IDE</b>	Entorno de desarrollo Integrado, utilizado para el desarrollo de software proveyendo funcionalidades al programador como detección de errores léxicos o sintácticos, formateo de código, compilar o ejecutar código, y entre los que destacan: Visual Studio Code, Netbeans, JetBrains.
<b>INSERT INTO</b>	Instrucción SQL.
<b>INT</b>	Tipo de dato SQL.
<b>Mb</b>	<i>Mega Bite</i> unidad de almacenamiento informático equivalente a 1000000 bytes.
<b>MySQL</b>	Es un DBMS de código abierto y el más popular en bases de datos relacionales, ofrece mejoras en velocidad, conectividad y seguridad.
<b>NOT NULL</b>	Instrucción SQL.
<b>Python</b>	Lenguaje de programación interpretado y a su vez multiparadigma, permitiendo aplicar el paradigma orientado a objetos, funcional o imperativa.

<b><i>RAM</i></b>	Del inglés (Random Access Memory), acceso de memoria aleatorio, es un almacenamiento volátil en donde la información que se almacena en ella es temporal y se elimina constantemente al ejecutar diferentes procesos en la computadora.
<b><i>SELECT</i></b>	Instrucción SQL.
<b><i>SET</i></b>	Instrucción SQL.
<b><i>SQL</i></b>	Del inglés (Structured Query Language), lenguaje de consulta estructurada utilizado para administrar una base de datos.
<b><i>TABLE</i></b>	Objeto de una base datos SQL.
<b><i>TIME</i></b>	Tipo de dato SQL.
<b><i>TRUNCATE</i></b>	Instrucción SQL.
<b><i>UPDATE</i></b>	Instrucción SQL.
<b><i>VARCHAR</i></b>	Tipo de dato SQL.
<b><i>VISUAL STUDIO CODE</i></b>	Entorno de desarrollo integrado.
<b><i>WHERE</i></b>	Instrucción SQL.



## RESUMEN

Actualmente CUNSUROC almacena sus registros estudiantiles en una base de datos de FoxBase Professional en su versión 2.5, dichos datos necesitan ser migrados a una nueva base de datos en producción MySQL, del Departamento de Procesamiento de Datos.

Para realizar esta migración se necesita realizar un proceso de ETL, en donde se necesitará extraer la información de los archivos (\*.dbf), esto con el apoyo de una librería en el lenguaje de programación Python, que permita realizar la transformación de datos en algunos registros, y finalmente generar archivos de inserción (\*.sql).

Estos archivos de inserción se estarán cargando inicialmente a una base de datos local, que representa una copia de la base de datos de producción; dicha base de datos local se encontrará corriendo en un contenedor de Docker basado en la imagen de MySQL, esto con el fin de reducir errores o inconsistencias al realizar la migración final a la base de datos de producción, y de esta manera realizar las correcciones que sean necesarias sin afectar directamente al ambiente productivo.

Las cargas se realizarán periódicamente para poder ir analizando que los datos registrados en cada tabla sean consistentes y de esta forma el trabajo sea más ordenado y la migración sea correcta.



## **OBJETIVOS**

### **General**

Realizar la correcta migración de la base de datos de CUNSUROC, al nuevo sistema de control académico.

### **Específicos**

1. Efectuar un análisis de los archivos registrados en la base de datos del CUNSUROC, para definir la estructura, contenido y magnitud de los mismos, y de esta manera planificar los procesos de migración de forma segmentada e incremental.
2. Definir las relaciones, restricciones y el esquema de la base de datos destino, mediante el acceso al modelo actual del Sistema de Control Académico para conservar la consistencia de los datos que serán migrados.
3. Implementar un proceso de ETL mediante scripts de programación para elaborar las lecturas, transformaciones y generación de los archivos de inserciones de los datos a migrar.
4. Ejecutar la migración mediante los archivos de inserción generados a una base de datos de pruebas a la cual se tendrá acceso y poder visualizar los datos migrados en el nuevo Sistema de Control Académico web.



## INTRODUCCIÓN

Las migraciones de bases de datos se emplean para conservar la información registrada de un sistema, para acoplarla a un nuevo sistema de información; muchas instituciones tanto públicas como privadas lo ven como un proceso que tarde o temprano lo deben cumplir, esto con el fin de mejorar y actualizar las tecnologías que utilizan para proveer un mejor producto o servicio a sus consumidores.

Existen diferentes categorías de migraciones de bases de datos, en las que se pueden mencionar Migraciones de almacenamiento: en las cuáles únicamente se cambian los dispositivos físicos de hardware en los que se almacenan la información por temas de rendimiento y eficiencia al registrar o consultar datos. Migraciones de aplicación: en la que se cambia la aplicación o sitio web que se utiliza para registrar la información y pueda que conlleve ciertos cambios a nivel de base de datos por temas de compatibilidad. Migraciones de bases de datos: en las que se cambian totalmente los proveedores y sistemas gestores de base de datos (DBMS), o se actualizan a una versión más reciente.

El proyecto actual implica una Migración de base de datos y a su vez de forma implícita se está realizando una Migración de aplicación; dado que actualmente el Centro Universitario Sur Occidente (CUNSUROC), utiliza una base de datos FoxBase Professional Versión 2.5 en donde está registrada la información de sus estudiantes en archivos con extensión (\*.dbf). Estos registros deben ser migrados a una base de datos MySQL del nuevo Sistema de Control Académico de la Universidad de San Carlos de Guatemala, el cual ofrece la capacidad de manejar diferentes funcionalidades al personal a cargo de control

académico de los diferentes centros regionales que lo utilizan, y debido a que la base de datos de origen como la de destino son totalmente diferentes; esta migración implica otro proceso y se explica a continuación:

La migración se realizará asemejando un proceso ETL (Extract, Transform and Load), por sus siglas en inglés, en el que se deberá extraer o leer los datos de los archivos (\*.dbf), realizar algunos procesos de transformación para adaptarlos a la estructura actual de las tablas de la base de datos destino y con ello estos datos se convertirán en archivos de inserción con extensión (\*.sql), para realizar la importación de estos hacia la base de datos destino.

Estos procesos se realizarán mediante un algoritmo programado utilizando librerías del lenguaje de programación Python para la generación y simulación del proceso ETL.

# 1. FASE DE INVESTIGACIÓN – CUNSUROC

A continuación, se presentará información acerca de esta Institución a la que se está apoyando para realizar el proceso de migración de registros estudiantiles.

## 1.1. Antecedentes de la institución

Se dará a conocer una breve reseña histórica, también la misión y visión que conforman esta institución.

### 1.1.1. Reseña histórica

El Centro Universitario de Suroccidente, fue creado por el Consejo Superior Universitario en el Acuerdo No. 372-77; de fecha 28 de septiembre de 1977, inició en febrero de 1978 con las carreras Técnico en Producción de Granos Básicos, Técnico en Procesamiento de Alimentos Vegetales y Técnico en Fruticultura Tropical. Actualmente cuenta con más de 15 carreras, en los que destacan distintos profesorados y licenciaturas en ramas como Administración de empresas, Psicología, ingeniería, entre otras.<sup>1</sup>

### 1.1.2. Misión

El Centro Universitario del Sur Occidente trabaja en equipo inter y multidisciplinario, formando profesionales competitivos, mediante la integración de programas de docencia, investigación y extensión; logrando estándares de calidad, reconocidos a nivel nacional e internacional, con la finalidad de promover el desarrollo de la región.<sup>2</sup>

---

<sup>1</sup> Universidad de San Carlos de Guatemala, CUNSUROC. *Información Cunsuroc: historia*. [informacioncunsuroc.blogspot.com](http://informacioncunsuroc.blogspot.com). Consulta: septiembre de 2021.

<sup>2</sup> Universidad de San Carlos de Guatemala, CUNSUROC. *Información Cunsuroc: misión y visión*. <http://informacioncunsuroc.blogspot.com/p/mision-y-vision.html>. Consulta: septiembre de 2021.

### **1.1.3. Visión**

Consolidar el liderazgo como centro de estudios superiores en la región del sur occidente, acreditado y reconocido a nivel nacional e internacional en la formación del recurso humano, altamente calificado, en distintas áreas del conocimiento, contribuyendo al desarrollo integral del país.<sup>3</sup>

### **1.1.4. Servicios que realiza**

Actualmente el Centro Universitario ofrece la oportunidad de estudiar diferentes carreras en plan diario, y plan fin de semana, entre las que destacan diferentes técnicos como: Procesamiento de Alimentos, Producción agrícola, Trabajo Social, Administración de Empresas, entre otros. También cuenta con Licenciaturas como: Ingeniería en Alimentos, Ingeniería en Agronomía Tropical, Ingeniería en Gestión Ambiental, Pedagogía y Administración Educativa, entre otras.

## **1.2. Descripción de las necesidades**

La actualización de sistemas es una necesidad que cualquier organización (pública o privada) necesitará conforme pasan los años, debido a que siempre existirán nuevas tecnologías que vienen ayudar a suplir muchas de las necesidades que surgen día a día, y con ello mejorar el rendimiento, optimizar procesos, con lo más valioso de las organizaciones; la información.

En el caso de CUNSUROC, necesitan actualizar la base de datos donde registran la información de sus estudiantes, que sería registrar las notas de los cursos aprobados al nuevo Sistema de Control Académico de la Universidad de San Carlos de Guatemala, el cuál contará con una aplicación web que será

---

<sup>3</sup> Universidad de San Carlos de Guatemala, CUNSUROC. *Información Cunsuroc: misión y visión*. <http://informacioncunsuroc.blogspot.com/p/mision-y-vision.html>. Consulta: septiembre de 2021.

administrable y accesible desde cualquier navegador y en cualquier dispositivo con acceso a internet.

La base de datos origen, se refiere a la base de datos actual del CUNSUROC utiliza Foxbase Proffesional versión 2.5, que actualmente almacena sus registros en archivos con extensión (\*.dbf), y se necesita migrar dicha información a una base de datos MySQL.

### **1.3. Priorización de las necesidades**

La migración de aproximadamente más de 250 000 registros estudiantiles de los cursos aprobados almacenados en la base de datos de origen que fue provista por el personal del CUNSUROC en el mes de febrero del 2022, dicha base de datos contiene información de más de 18 carreras con grados Técnicos y Licenciaturas, éstas cuentan con información de todos sus cursos aprobados aproximadamente desde el año 1983, en diferentes períodos de cada semestre como notas de exámenes finales, primer y segunda retrasada, equivalencias, entre otros.

El acceso eficaz a los datos almacenados para mejorar la rapidez de los procesos internos del centro universitario, como generación de certificados de cursos aprobados, asignación de cursos, consulta de notas de cursos asignados y aprobados, asignación de cursos por semestre, entre otras funcionalidades que provee el nuevo Sistema de Control Académico.



## **2. FASE TÉCNICO PROFESIONAL**

Se realizará una migración de registros estudiantiles pero que a su vez incluye una migración de aplicación ya que los registros se encuentran en un sistema antiguo y pasaran a formar parte del nuevo Sistema de Control Académico en un gestor de base de datos relacional más nuevo.

### **2.1. Descripción del proyecto**

El proyecto actual contempla una migración un tanto diferente, a causa de que, al ser una base de datos de origen no relacional, y que también guarda sus registros en archivos planos, se podría utilizar una herramienta o bien personalizar un algoritmo en algún lenguaje de programación para realizar esta tarea.

Para darse una idea de lo que se habla, estos archivos se deben leer extrayéndoles su información y realizar relaciones, pero únicamente de forma lógica; es decir que se maneja mediante comparación de códigos que simulan las llaves primarias que se conocen en una base de datos relacionales, dado que no se cuenta con la opción de ejecutar consultas, y con ello se logra simular un proceso de extracción para continuar con la transformación de los datos; en caso se necesite separar códigos que se encuentren en un solo registro para identificar cursos, carnets, entre otros.

Y para todo esto se utilizará Python como lenguaje de programación, para ejecutar el proceso de lectura y transformación de estos registros, para

finalmente generar scripts SQL para poder realizar la inserción a la base de datos final.

Se cuenta también con acceso a la base de datos destino a través de internet con credenciales propias, dicho acceso fue proveído por Procesamiento de Datos, para facilitar el proceso de la migración de forma remota, y con ello poder continuar con la migración de forma remota.

## **2.2. Sistemas de bases de datos**

Un sistema de bases de datos es software que se utiliza para registrar información viéndose como un tipo de armario electrónico para archivar datos, en la que los usuarios pueden realizar diferentes operaciones:

- Crear nuevos registros
- Agregar registros a archivos existentes
- Recuperar registros
- Modificar o actualizar registros
- Eliminar registros
- Eliminar archivos de bases de datos

Las bases de datos están conformadas básicamente por 4 componentes: datos o información, hardware (discos duros o de estado sólido, procesadores, memoria principal), donde se guardan física y lógicamente los datos, software (sistema para acceder a los datos DBMS), y usuarios (administradores de bases de datos, programadores y quienes visualizan los datos los usuarios finales).

Se conoce como base de datos ya que sus datos son persistentes lógicamente en el transcurso y flujo normal, hasta que exista una alteración de

parte de los usuarios o problemas externos con el hardware o software gestor de la base de datos.

Las bases de datos representan el modelo de negocio de las empresas u organizaciones públicas o privadas, simulando los flujos de los procesos internos en cada área de trabajo, y permitan llevar el registro de estos procesos con el fin de aprovechar estos datos para fines estratégicos, estadísticos, o de mejora continua.

### **2.2.1. Entidades y vínculos**

Representan las entidades en las bases de datos y define los vínculos que pueden existir entre ellas.

#### **2.2.1.1. Entidad**

Se refiere a algo tangible, o no dentro de una empresa u organización, por ejemplo: empleados, áreas de trabajo, puesto, entre otros.

#### **2.2.1.2. Vínculo**

Los vínculos son las relaciones que existen dentro de las entidades por ejemplo un departamento tiene un vínculo con la entidad empleado, debido la misma cuenta con 1 o varios empleados, el vínculo permite obtener consultas referentes a estas entidades; como saber el número de empleados que trabajan en el departamento de informática, cuantas áreas de trabajo existen por departamento, entre otros.

## **2.2.2. Propiedades**

Conforman la información que se quiere registrar de las entidades, por ejemplo, la información de la entidad Empleado podría ser: nombre, fecha de nacimiento, número de identificación, salario, departamento, entre otros.

## **2.2.3. Datos y modelos de datos**

Se procede con la definición de que son los datos y la representación de estos dentro de un modelo de datos.

### **2.2.3.1. Datos**

“La palabra datos deriva del latín ‘dar’, por eso se dice que los datos son hechos dados, y de ellos se puede inferir hechos adicionales o proposiciones, y por ende una base de datos contiene una colección de proposiciones verdaderas”.<sup>4</sup>

### **2.2.3.2. Modelo de datos**

Es una definición lógica que representa la abstracción de los objetos y operadores que conforman el sistema o modelo de negocio deseado, los objetos permiten modelar la estructura de datos y los operadores: su comportamiento; este modelo conlleva una implementación que es la transformación de este en algo real, el modelo es algo que los usuarios pueden conocer, pero la implementación no.

---

<sup>4</sup> DATE, C.J. *Introducción a las bases de datos*. p. 13.

## **2.2.4. SQL**

(Structured query language), lenguaje de consulta estructurada utilizado para manipular sistemas relacionales de bases de datos. Se basa en el manejo del álgebra y el cálculo relacional, realizando operaciones como unión, intersección, de la teoría de conjuntos aplicados a sentencias como UNION, JOINS.

### **2.2.4.1. Tipos de datos**

Estos dependerán del motor de base de datos instalado, pero los tipos más comunes son:

- Varchar: recibe una cadena de caracteres limitada por la cantidad definida en su declaración.
- Int: recibe valores enteros, con o sin signo.
- Decimal: recibe valores decimales, con el número de decimales definidos en su declaración.
- Date: permite valores tipo fecha en el formato regular de 'YYYY-MM-DD'.
- Time: permite registra la hora del día en horas minutos segundos 'hh:mm:ss'

#### **2.2.4.1.1. DDL**

Lenguaje de definición de datos (Data Definition Language): son las sentencias declarativas para crear objetos en la base de datos, en sql se refiere a las sentencias:

- Create: permite crear objetos como tablas, vistas, procedimientos almacenados, trigger, funciones, índices, entre otros.

Ejemplo de creación de una tabla:

```
CREATE TABLE alumnos;
```

- Alter: permite modificar la estructura de tablas u objetos, para agregar nuevos campos a una tabla, modificar algún tipo, agregar/eliminar índices, modificar un procedimiento/trigger.

Ejemplo de una instrucción para agregar una columna a una tabla:

```
ALTER TABLE alumnos ADD edad INT NOT NULL;
```

- Drop: elimina un objeto (tabla, vista, entre otros), permite utilizar cláusula where.

Ejemplo de eliminación de registros con condición:

```
DROP TABLE alumnos WHERE edad > 20;
```

- Truncate: permite eliminar el contenido completo de una tabla, no permite el uso del comando where.

Ejemplo de eliminación de registros de una tabla:

```
TRUNCATE TABLE alumnos.
```

#### 2.2.4.2. DML

Lenguaje de manipulación de datos (Data manipulation language): permite realizar consultas, insertar datos, actualizar datos a los usuarios sobre las tablas y los registros en ella, aplicando diferentes cláusulas para realizar diferentes filtros de datos.

- Select: permite obtener registros de una o más tablas, aplicando diferentes criterios del algebra relacional como unión, intersección, entre otros.

Ejemplo de una consulta a dos tablas relacionándolas por las llaves foráneas y primarias:

```
SELECT e.nombre, e.edad, c.curso FROM Estudiante e,  
curso_asignacion c, WHERE c.estudiante_carnet = e.carnet
```

- Insert: esta instrucción permite registrar uno o múltiples registros a una tabla.

Ejemplo de instrucción para inserción de registros en una tabla:

```
INSER INTO my_table (name_field1, [name_fieldN]*)? VALUES  
(value1, [valueN]*)
```

- Update: permite actualizar uno o múltiples registros dependiendo si cumplen o no ciertas condiciones.

Ejemplo de actualización de uno o más registros que cumplan con la condición:

```
UPDATE my_table SET field1 = 'new value' WHERE field2 =  
'condition_value'
```

## **2.2.5. Migración de bases de datos**

Las migraciones de bases de datos dependerán mucho del ambiente y circunstancias en las que se encuentren los datos, es por ello que se puede clasificar algunas de ellas.

### **2.2.5.1. Categorías**

Existen diferentes categorías de bases de datos, pero se destacan las siguientes:

#### **2.2.5.1.1. Migración de base de datos**

Se dan cuando toda la información registrada pasará a un nuevo modelo de datos, ya sea de un sistema gestor de base de datos igual o a otro totalmente diferente, para realizar este tipo de migraciones, existen procesos que se pueden aplicar como los ETL (Extract, Transform and Load), apoyado de herramientas o bien en base a algoritmos en algún lenguaje de programación, para simular todas las fases de este, que se detallan más adelante.

#### **2.2.5.1.2. Migración de aplicación**

Sucede cuando se quiere actualizar totalmente un sistema, en los que las tecnologías que se utilizarán cambiarán totalmente, ya sea el lenguaje de programación del sitio web o aplicación de escritorio, y de igual forma la base de

datos que se utilizará, por ejemplo, de pasar de guardar en archivos de texto planos a un modelo relacional, o a bases de datos no relacionales.

### **2.2.5.1.3. Migración de almacenamiento**

Este tipo de migración se da cuando se quiere aumentar la capacidad del servidor con nuevos dispositivos hardware con mayor rendimiento, espacio o velocidad de transferencia de datos, regularmente cambio de discos duros por discos de estado sólido, o bien aumentando la memoria principal RAM.

### **2.2.6. FoxBase Professional versión 2.5**

FoxBase es un lenguaje de programación orientado a procedimientos, y también un sistema gestor de bases de datos.

En el año 1993 fue lanzada esta versión para computadoras con sistema operativo Windows, y también la primera con soporte de interfaz gráfica, para junio de ese mismo año se lanza la subversión 2.5a con soporte de idiomas internacionales; para septiembre fue lanzada la subversión 2.5b con soporte ahora para griego, checo, polaco e islandés.

Esta versión es la que se utiliza en el Centro Universitario Sur Occidente, este tipo de base de datos registran sus datos en archivos con extensión \*.dbf, y a decir verdad no deberían ser consideradas como bases de datos, a pesar de contar con relaciones entre tablas, no permite la transaccionalidad que es un principio fundamental de las bases de datos.

### **2.2.7. MySQL**

MySQL es un sistema de administración de base de datos (DMBS por sus siglas en inglés), y la más popular de código abierto, siendo distribuida, desarrollada y soportada por Oracle Corporation.

Una base de datos es una colección de datos de forma estructurada, y para poder administrar esta información que puede ir desde algo tan simple como una lista de estudiantes; hasta la información registrada en un clip, audio o video. Para poder acceder, actualizar o eliminar estos se necesita de un sistema administrador como lo es MySQL, dicho sistema se basa en la teoría relacional y en SQL (Structured query Language), que es el lenguaje que permitirá manipular todos los datos basados en ANSI/ISO SQL Standard y a la vez es un estándar internacional que define la estructura del lenguaje y ciertas reglas que deben cumplir.

MySQL es un DBMS rápido, ofrece muchas mejoras como conectividad, velocidad y seguridad.

### **2.2.8. ETL**

Debido a que muchas veces las organizaciones desean utilizar nuevas tecnologías y conservar sus datos, limpiar o eliminar registros duplicados, hace esta necesidad que existan estos procesos ETL (Extract, Transform and Load por sus siglas en Inglés), este es un proceso que sigue un flujo en el que permite la carga de diversas fuentes de información, en la que a través de un motor: carga esta información a tablas temporales de las que posteriormente se realizarán las diversas transformaciones (eliminaciones, actualización o creación), de nuevos campos en los registros, para posteriormente poder ser cargado a una base de

datos destino. Cabe destacar que este es un proceso un poco más complejo de lo que parece, y que existen herramientas que ayudan a facilitar estos procesos, pero no es la única forma de simular un ETL, también se puede realizar mediante un algoritmo de programación, como la solución que seguidamente se presenta.

### **2.3. Solución del proyecto**

Mediante un algoritmo en Python se realizará la fase de extracción y transformación de los datos de la base de datos origen representados en archivos \*.dbf, con el tratamiento correcto de estos datos se obtendrán archivos \*.sql con sentencias de inserción a las tablas de la base de destino implicadas.

La base de datos origen se encuentra organizada por un archivo que contiene la información de los pensum con todos los cursos que se imparten en cada carrera, para obtener la información de la carrera se creó un archivo \*.csv, con la información de cada carrera: código, nombre, plan, extensión, correlativo, y el grado (técnico o licenciatura), y la base de datos de origen contiene carpetas. Cada carpeta es nombrada con el correlativo de cada carrera, por ejemplo, la carrera: Técnico en Administración de Empresas se representa con el correlativo "ADM", en cada carpeta existe un archivo principal que contiene el registro de todas las actas, estas actas a su vez representan un archivo dentro de esta carpeta, dichas actas contienen los registros estudiantiles de las notas de cursos aprobados que deberán ser migrados.

Se definió de forma explícita la configuración de que carpetas correspondientes a cada carrera deben ser migradas, también los archivos principales de cada una, esto con el fin de que el algoritmo entienda en donde leer la información.

Las actas que contienen la información de los registros estudiantiles cuentan con la siguiente estructura: fecha, código (contiene el código y nombre del curso), carnet, nombre (nombre de estudiante), zona, fin, prom (nota final), en base a este archivo se deberá realizar la extracción de la información del estudiante, del curso, y la nota del curso aprobado, es en estos archivos en donde se realizarán la mayoría de las transformaciones necesarias.

Tabla I. Estructura de un acta

	A	B	C	D	E	F	G	H
1	FECHA	CODIGO	CARNET	NOMBRE	ZO	FIN	PROM	
2	29/06/1990	002 MATEMATICA I	8730810	SANDOVAL, OSCAR E.	71	71		
3	29/06/1990	002 MATEMATICA I	8840108	MARTINEZ FLORES, HOMERO	64	64		
4	29/06/1990	002 MATEMATICA I	8840169	VELASQUEZ M., JORGE F.	65	65		
5	29/06/1990	002 MATEMATICA I	8840238	RAYMUNDO RAYMUNDO, LEONEL	74	74		
6	29/06/1990	002 MATEMATICA I	8940252	RIVERA PEDROZA, ARIEL F.	60	60		
7	29/06/1990	002 MATEMATICA I	9040406	ALVAREZ VASQUEZ, RODOLFO	75	75		
8	29/06/1990	002 MATEMATICA I	9040407	SANDOVAL ALVAREZ, FELIPE	66	66		
9	29/06/1990	002 MATEMATICA I	9040418	MENDEZ GONZALEZ, SERGIO	72	72		
10	29/06/1990	002 MATEMATICA I	9040425	PELAEZ CACERES, LUDIM H.	73	73		
11	29/06/1990	002 MATEMATICA I	9040468	BATRES ESTRADA, ROLANDO	68	68		
12	29/06/1990	002 MATEMATICA I	9040479	VELASQUEZ SACAYON, JUAN	73	73		
13	29/06/1990	002 MATEMATICA I	9040483	LUARCA GONZALEZ, KARLA	81	81		
14	29/06/1990	002 MATEMATICA I	9040514	MARROQUIN MARTINEZ, ALMA	62	62		
15	29/06/1990	002 MATEMATICA I	9040969	ROMERO VAHLE IIIIS A	74	74		

Fuente: Universidad de San Carlos de Guatemala, Centro Universitario de SuroOccidente, CUNSUROC. Acta estudiantil del curso de Matemática referente al periodo de vacaciones del año 1990. 002VAC90.dbf abierto en Excel.

El proyecto se realizó en el lenguaje Python utilizando el editor de texto Visual Studio Code creando un repositorio en Github para el control de versiones de este, el proyecto se encuentra publicado en el siguiente enlace: <https://github.com/jefflinares/CUNSUROC> con acceso privado, si desea obtener acceso al mismo enviarme un correo a [jefflinaresceron@gmail.com](mailto:jefflinaresceron@gmail.com).

Para la lectura de los archivos DBF se utilizó la librería dbfread, se encuentra disponible para instalar a través de pip y requiere una versión de Python no menor a 3.2 o 2.7, solo basta ejecutar el siguiente comando en la consola:

`“pip install dbfread”`.<sup>5</sup>

La estructura del proyecto en Python se definió de la siguiente forma, un archivo principal, una carpeta llamada “clases” la cual contiene todas las clases que se utilizaron, una carpeta “utils” que contiene el algoritmo ETL, y en esta existe una carpeta llamada “archivosGenerados” donde se generaron los archivos \*.sql y archivos \*.json con información de errores generados y también archivos de información de estudiantes.

Luego de todo el proceso de extracción y transformación se utilizará una base de datos local para la carga de los archivos de inserción, esta base se encontrará dentro de un contenedor de Docker creado a partir de la imagen de MySQL y se utilizará el cliente de PHPMysql para acceder a la base de datos desde el navegador, se realizará este proceso para corregir cualquier inconveniente previo a realizar la carga a la base de datos de pruebas y posteriormente a la base de datos de producción, de esta manera el margen de error e inconsistencia de datos en el ambiente productivo se verá disminuido en gran medida.

Se ampliará más acerca de la implementación de este algoritmo, iniciando por el entorno de desarrollo local definido en el apéndice 1 y la explicación del algoritmo ETL en el apéndice 2.

---

<sup>5</sup> Pypi org. *Instalación de librería dbfread*. <https://pypi.org/project/dbfread/>. Consulta: septiembre de 2021.

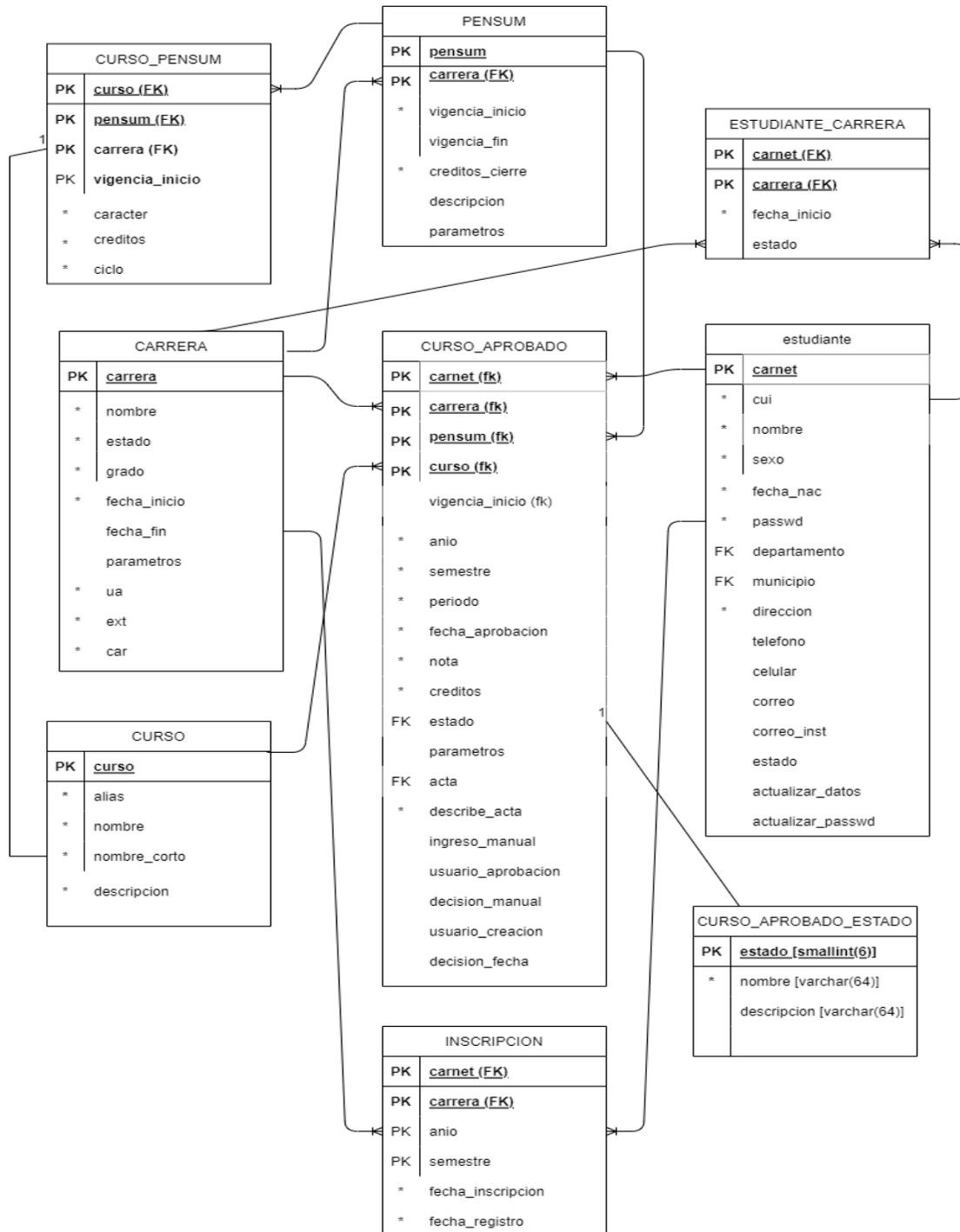
### 2.3.1. Base de datos destino

El sistema de base de datos destino proveído por la Unidad de Procesamiento de Datos se visualizará en el siguiente diagrama relacional, dentro de las tablas con información migrada resaltan:

- Carrera: cuenta con el registro de las carreras con información como código, nombre, vigencia de inicio, unidad académica, extensión.
- Pensum: contiene la información de todos los pensums dependiendo de las carreras entre sus campos se encuentran: carrera, vigencia de inicio.
- Curso: se encuentran registrados los cursos impartidos en cada carrera, entre sus campos se encuentra: código curso, y el nombre.
- Curso\_pensum: esta tabla se encarga de relacionar cada curso por carrera y por pensum.
- Estudiante: se registra información del estudiante como carnet, nombre, cui, sexo, entre otros.
- Curso\_aprobado: en esta tabla es en donde registradas todas las notas que se leyeron, entre sus campos se encuentran carnet, carrera, pensum, curso, vigencia\_inicio, anio, semestre, periodo, fecha\_aprobacion, nota, entre otros.

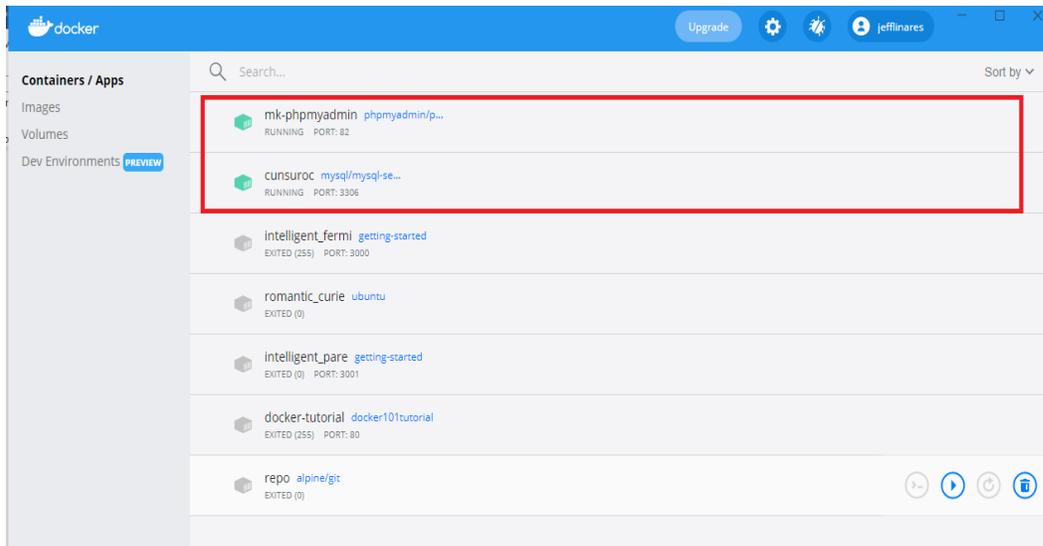
En la siguiente figura se presenta el diagrama de base de datos extraído lógicamente.

Figura 1. Diagrama de Modelo Relacional de la base de datos destino



Fuente: elaboración propia, empleando draw.io.

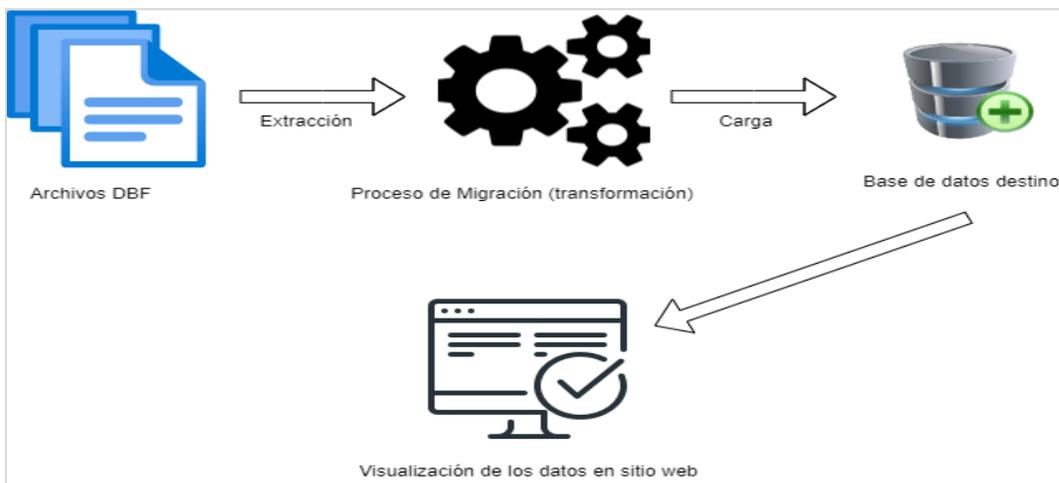
Figura 2. Contenedores de base de datos local



Fuente: elaboración propia, empleando captura de pantalla del programa Docker Desktop 2021.

En el siguiente diagrama podrá visualizar de mejor forma el proceso de migración de forma general.

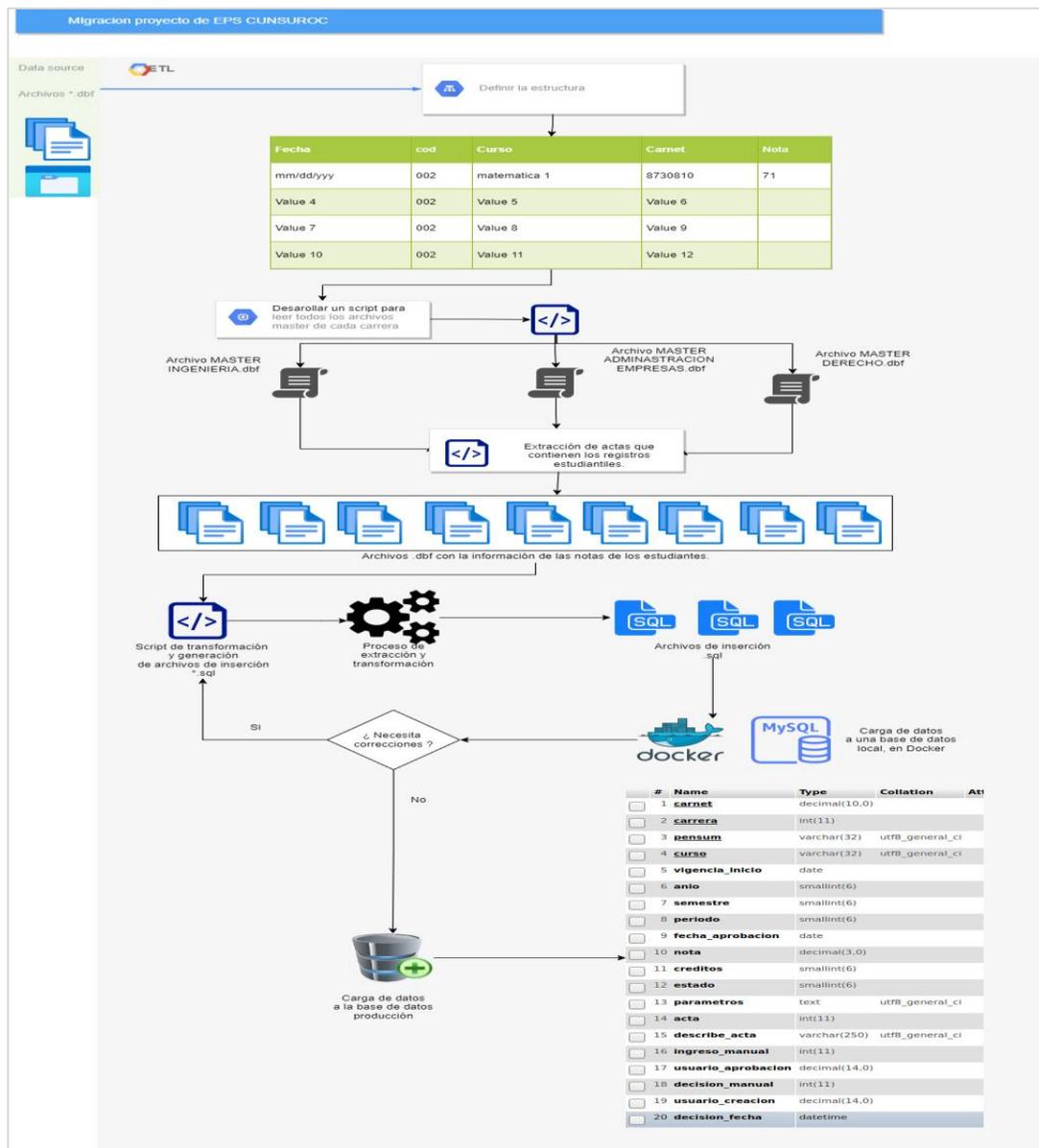
Figura 3. Diagrama de flujo de proceso ETL general



Fuente: elaboración propia, empleando draw.io.

Y el siguiente diagrama es más específico, detalla el proceso de extracción, lectura y transformación y la generación de archivos de inserción.

Figura 4. Diagrama de flujo del proceso de migración



Fuente: elaboración propia, empleando draw.io.

### 2.3.2. Información migrada

Finalmente, luego de haber realizado el proceso ETL y obtener los archivos de migración e insertarlos en la base de datos local utilizando Docker, y en ella haber corregido algunos inconvenientes e inconsistencias presentadas; se procedió a cargar los datos a una base de datos de pruebas, que a partir de la misma se migrará la información a la base de datos de producción del nuevo Sistema de Control Académico.

En la siguiente tabla, se presenta información cuantitativa de los registros migrados, asociados a las tablas correspondientes.

Tabla II. **Resumen de registros migrados**

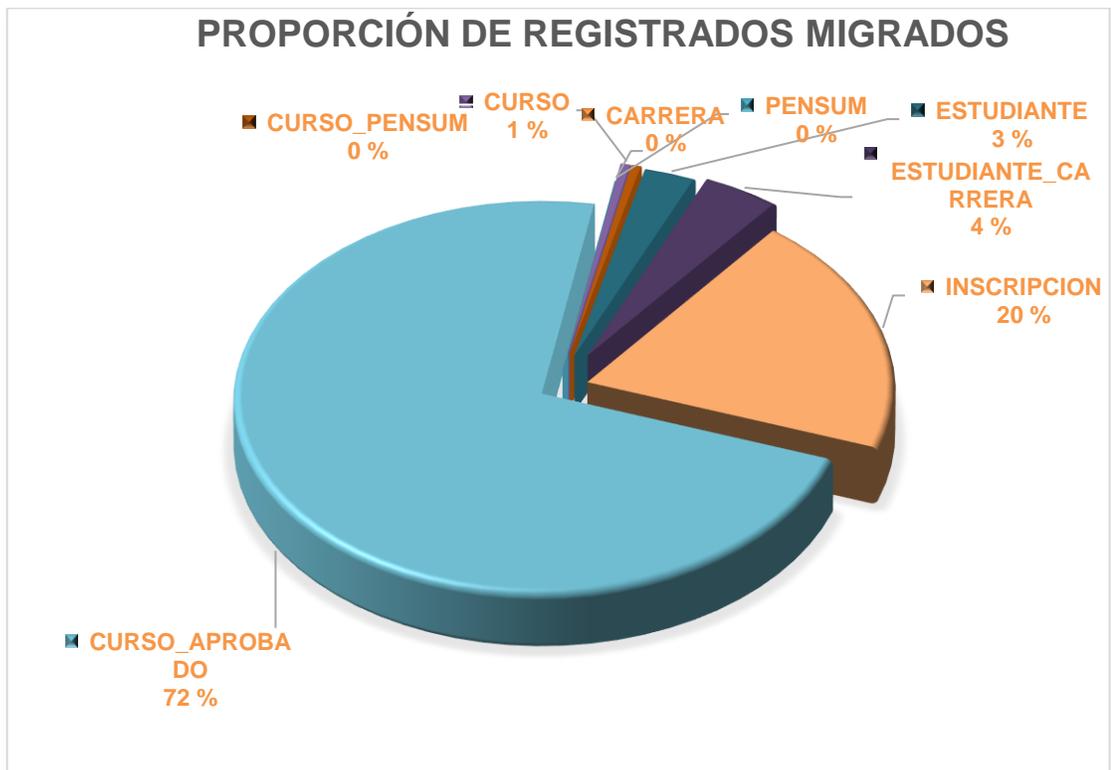
TABLA	REGISTRADOS MIGRADOS
CARRERA	22
PENSUM	36
CURSO	1 940
CURSO_PENSUM	1 932
ESTUDIANTE	10 556
ESTUDIANTE_CARRERA	15 964
INSCRIPCION	77 145
CURSO_APROBADO	281 734
TOTAL REGISTROS	389 329

Fuente: elaboración propia.

El registro de más de 15 mil estudiantes y más de 280 mil registros de cursos aprobados fueron migrados de información registrada desde el año 1985 hasta diciembre del año 2021.

En la siguiente gráfica se observa la proporción del total de registros migrados por tabla.

Figura 5. Proporción de registros migrados



Fuente: elaboración propia.

Los cursos aprobados representan más del 72 % de la información migrada.

El proceso de ejecución del algoritmo ETL tardó aproximadamente 35 minutos, en leer la información de las carreras, procesar, transformar y generar los archivos de inserción SQL.

### 2.3.3. Sistema de control académico WEB

Este sistema permitirá tanto a personal administrativo como a estudiantes del CUNSUROC a poder gestionar la información referente a gestión de carreras, pensum, cursos, consultas entre otras funcionalidades para el personal administrativo; consultas de cursos aprobados y asignados para estudiantes.

El acceso al sistema será por medio de credenciales proveídas por la Unidad de Procesamiento de Datos; se obtuvo acceso a dicho sistema con un rol de administrador en el ambiente de pruebas para verificar la información migrada, y se mostrarán algunas de las consultas realizadas utilizando el usuario con rol de administrador.

Figura 6. Consulta de cursos aprobados por estudiante y carrera

Año	Semestre	Carrera	Codigo Curso	Curso	Sección	Periodo	Zona	Final	Nota Asignación	1ra. Recuperación	2da. Recuperación	Fecha de Aprobación
2008	1	Trabajo Social	137-04-31	ANTROPOLOGIA GENERAL	Equivalencia	Semestre regular	0.00	89	89	---	---	2008-05-15
2008	1	Trabajo Social	529-04-376	TEORIA POLITICA	Equivalencia	Semestre regular	0.00	71	71	---	---	2008-05-12
2008	1	Trabajo Social	546-04-371	PLANIFICACION SOCIAL	Equivalencia	Semestre regular	0.00	82	82	---	---	2008-05-13
2008	1	Trabajo Social	753-04-712	PRACTICA DE INVESTIGACION SOCIAL I	Equivalencia	Semestre regular	0.00	75	75	---	---	2008-05-19
2008	1	Trabajo Social	751-04-695	METODOLOGIA DE INVEST SOCIAL I	Equivalencia	Semestre regular	0.00	70	70	---	---	2008-05-14
2008	2	Trabajo Social	531-04-368	PSICOLOGIA GENERAL	Equivalencia	Semestre regular	0.00	64	64	---	---	2008-07-11
2007	1	Trabajo Social	138-04-19	FILOSOFIA GENERAL	Equivalencia	Semestre regular	0.00	80	80	---	---	2007-05-15
2007	1	Trabajo	720-04-	INTRODUCCION AL	Equivalencia	Semestre	0.00	72	72	---	---	2007-05-14

Fuente: Universidad de San Carlos de Guatemala, Procesamiento de Datos, Sistema de Control Académico. *Cursos aprobados.*

[https://epsdpd.usac.edu.gt/eps\\_cunsuroc/Administrativo/cursosAprobados](https://epsdpd.usac.edu.gt/eps_cunsuroc/Administrativo/cursosAprobados). Consulta: marzo de 2022.

Figura 7. Consulta de cursos asignados por estudiante

**CURSOS ASIGNADOS - BUSCAR**

Registro Académico

**Cursos Asignados**  
 Estudiante: ██████████ - ██████████

Año	Semestre	Carrera	Codigo Curso	Curso	Sección	Periodo	Zona	Final	Nota Asignación	1ra. Recuperación	2da. Recuperación
2008	1	Trabajo Social	137-04-31	ANTROPOLOGIA GENERAL	Equivalencia	Semestre regular	0.00	89	89	---	---
2008	1	Trabajo Social	529-04-376	TEORIA POLITICA	Equivalencia	Semestre regular	0.00	71	71	---	---
2008	1	Trabajo Social	546-04-371	PLANIFICACION SOCIAL	Equivalencia	Semestre regular	0.00	82	82	---	---
2008	1	Trabajo Social	753-04-712	PRACTICA DE INVESTIGACION SOCIAL I	Equivalencia	Semestre regular	0.00	75	75	---	---
2008	1	Trabajo Social	751-04-695	METODOLOGIA DE INVEST. SOCIAL I	Equivalencia	Semestre regular	0.00	70	70	---	---
2008	2	Trabajo Social	531-04-368	PSICOLOGIA GENERAL	Equivalencia	Semestre regular	0.00	64	64	---	---
2007	1	Trabajo Social	138-04-19	FILOSOFIA GENERAL	Equivalencia	Semestre regular	0.00	80	80	---	---
2007	1	Trabajo Social	720-04-672	INTRODUCCION AL TRABAJO SOCIAL	Equivalencia	Semestre regular	0.00	72	72	---	---
2007	1	Trabajo Social	721-04-673	INVEST.BIBLIOGRAFICA Y DOCUMENTAL	Equivalencia	Semestre regular	0.00	74	74	---	---

Fuente: Universidad de San Carlos de Guatemala, Procesamiento de Datos, Sistema de Control Académico. *Cursos asignados.*

[https://epsdpd.usac.edu.gt/eps\\_cunsuroc/Administrativo/cursosAsignados](https://epsdpd.usac.edu.gt/eps_cunsuroc/Administrativo/cursosAsignados). Consulta: marzo de 2022.

Los estudiantes tendrán acceso a vistas similares para consultar su información, como por ejemplo una vista de cursos aprobados, pero a su vez se desplegarán las carreras en las que se encuentran asignados.

Figura 8. Vista de cursos aprobados por el estudiante y carrera

**CURSOS APROBADOS - BUSCAR**

Carrera: 220014 - Licenciatura en Trabajo Social

**Cursos Aprobados**  
Estudiante: [REDACTED] - [REDACTED]

Año	Semestre	Carrera	Codigo Curso	Curso	Sección	Periodo	Zona	Final	Nota Asignación	1ra. Recuperación	2da. Recuperación	Fecha de Aprobación
2020	1	Licenciatura en Trabajo Social	547	FORMUL Y EVAL DE PROY. SOCIALES	Equivalencia	Semestre regular	0.00	63	63	---	---	2020-05-18
2020	1	Licenciatura en Trabajo Social	741	CIENCIA Y EPISTEMOLOGIA	Equivalencia	Semestre regular	0.00	61	61	---	---	2020-05-13
2020	1	Licenciatura en Trabajo Social	743	PROMOCION SOCIAL Y EDUCACION POPULAR	Equivalencia	Semestre regular	0.00	80	80	---	---	2020-05-21
2020	1	Licenciatura en Trabajo Social	755	SEMINARIO DE ORGANIZACION SOCIAL	Equivalencia	Semestre regular	0.00	80	80	---	---	2020-05-21
2020	2	Licenciatura en Trabajo Social	756	SEMINARIO DE POLITICAS SOCIALES	Equivalencia	Semestre regular	0.00	65	65	---	---	2020-11-20

Total de Cursos Aprobados: 5

© 2017 - 2022 by Universidad de San Carlos de Guatemala. All rights reserved.

Fuente: Universidad de San Carlos de Guatemala, Unidad de Procesamiento de Datos, Sistema de Control Académico. *Cursos aprobados desde la vista de estudiante.*

[https://epsdpd.usac.edu.gt/eps\\_cunsuroc/Estudiante/cursosAprobados](https://epsdpd.usac.edu.gt/eps_cunsuroc/Estudiante/cursosAprobados). Consulta: marzo de 2022.

## 2.4. Costos del proyecto

Entre los costos del proyecto destacan recursos humanos y materiales, se incluyen los servicios de hogar necesarios para trabajar desde casa.

Tabla III. **Costos asociados al proyecto**

Recurso	Rol	Cantidad	Unidad	Costo Unitario	Subtotal
Jefferson Linares	Desarrollador de software	560	Horas	Q120,00	Q67 200,00
Ing. Herman Veliz	Asesor de escuela	20	Horas	Q250,00	Q5 000,00
Lic. Rodrigo Mendizábal	Asesor de proyecto	25	Horas	Q250,00	Q6 250,00
Estuardo Tello	Asesor técnico	15	Horas	Q150,00	Q2 250,00
Laptop	Equipo de cómputo	1	Laptop	Q7 500,00	Q7 500,00
Internet residencial	Servicios	7	Meses	Q150,00	Q1 050,00
Servicios básicos en el hogar	Servicios	7	Meses	Q1 500,00	Q10 500,00

Total proyecto	Q99 750,00
----------------	------------

Fuente: elaboración propia.

## 2.5. Beneficios del proyecto

Entre los beneficios del proceso de migración realizado destacan los siguientes:

- Los miembros de control académico del CUNSUROC dejarán de utilizar un sistema antiguo y robusto para el registro de información estudiantil, porque les limitaba ciertas funcionalidades, en su lugar utilizarán el Sistema de Control Académico WEB ya existente proveído por la Unidad de Procesamiento de datos, en donde el acceso al mismo será mediante credenciales con usuarios de rol administrativo, proveídos a personal calificado de este centro regional, este sistema web consume la base de datos a la cual se migró la información que existía en la base de datos de origen y ahora dicha información es visible en este sistema.
- Entre las funcionalidades que tendrán miembros de control académico del CUNSUROC en este nuevo sistema destacan:

- Gestión de carreras, pensum, cursos, docentes; usuarios; estudiantes; salones; períodos; horarios; actas, entre otros.
- Consultas como cursos asignados y aprobados por estudiante.
- Reportes como estudiantes por carrera o de equivalencias.

Cabe resaltar que este sistema ya existente no fue modificado, mucho menos alterado en cualquiera de sus funcionalidades durante el desarrollo de este proceso de migración, únicamente se hace énfasis de algunas funcionalidades que se podrán realizar.

- El centro regional seguirá creciendo tanto en registro de estudiantes como también, en desarrollo tecnológico al empezar a utilizar herramientas más actualizadas, rápidas y fiables.
- Los estudiantes del CUNSUROC también serán beneficiados con esta migración, ahora tendrán la posibilidad de ver su información de los cursos aprobados en el mismo sistema web mencionado; mediante credenciales de usuario y contraseña, ellos contarán con algunas de las funcionalidades siguientes:
  - Consultar notas de cursos asignados y aprobados
  - Gestiones como solicitud de certificaciones: las mismas se podrían realizar de una forma más rápida, y a su vez les permitirá llevar un seguimiento de éstas.

- Los profesores también tendrán acceso a este sistema web, contarán con algunas funcionalidades como: gestión de cursos impartidos; control de notas, entre otras.



### **3. FASE DE CAPACITACIÓN**

La fase de capacitación será impartida principalmente por la Unidad de Procesamiento de Datos al personal de Control Académico del CUNSUROC, en la utilización del nuevo Sistema de Control Académico, y también se brindará una pequeña capacitación acerca del funcionamiento del algoritmo implementado en Python, la forma de ejecutarse y configurarse para la lectura de los archivos de entrada.

#### **3.1. Capacitación propuesta del Sistema de Control Académico**

La Unidad de Procesamiento de Datos impartirá la capacitación de su sistema por medio de videoconferencias, que se llevarán a cabo en diferentes sesiones, el tiempo que se considere necesario, para ello se utilizarán usuarios administradores en el ambiente de pruebas del Sistema de Control Académico, con el fin de que el personal de Control Académico se sienta en la libertad de agregar, modificar y eliminar los datos registrados en la base de datos, sin afectar información que se encuentre en el ambiente de producción; de esta manera se podrán familiarizar de una forma mucho más rápida a este sistema, conociendo las funcionalidades para gestionar carreras, pensum, cursos, estudiantes, reportes, cualquier funcionalidad que el sistema les ofrece.

##### **3.1.1. Manual de usuario**

Se documentará de forma breve el uso de las principales funciones del sistema web, cabe destacar que dicho manual no será oficial, pero servirá de referencia para disposición del personal de Control Académico en caso necesiten

capacitar a nuevo personal. La estructura del manual será muy básica y se detalla a continuación:

- Descripción del sistema
- Funcionalidades administrativas
  - Inicio de sesión
  - Consultas
    - Cursos asignados
    - Cursos aprobados
  - Carreras
    - Gestión de carreras
    - Gestión de pensum
    - Gestión de cursos
  - Reportes
- Funcionalidades estudiantiles
  - Consultas
    - Cursos asignados
    - Cursos aprobados
  - Asignación de cursos
  - Gestiones

- Solicitud de certificación
- Perfil
  - Cambio de contraseña

### **3.2. Capacitación algoritmo ETL**

Esta capacitación será más técnica y realmente se dará a conocer el flujo del algoritmo, la configuración necesaria en el ambiente local para poder ejecutar dicho algoritmo, aunque difícilmente será utilizado de nuevo por el personal de Control Académico porque el mismo se utilizó para generar los archivos \*.sql y debería ser ejecutado preferiblemente por alguien con conocimientos informáticos para facilitar la ejecución del mismo, aunque cabe resaltar que la base de datos de origen dejará de ser utilizada, por tal motivo no habría más datos por migrar a la base de datos del Sistema de Control Académico web, que será la que se esté utilizando para registrar nueva información.

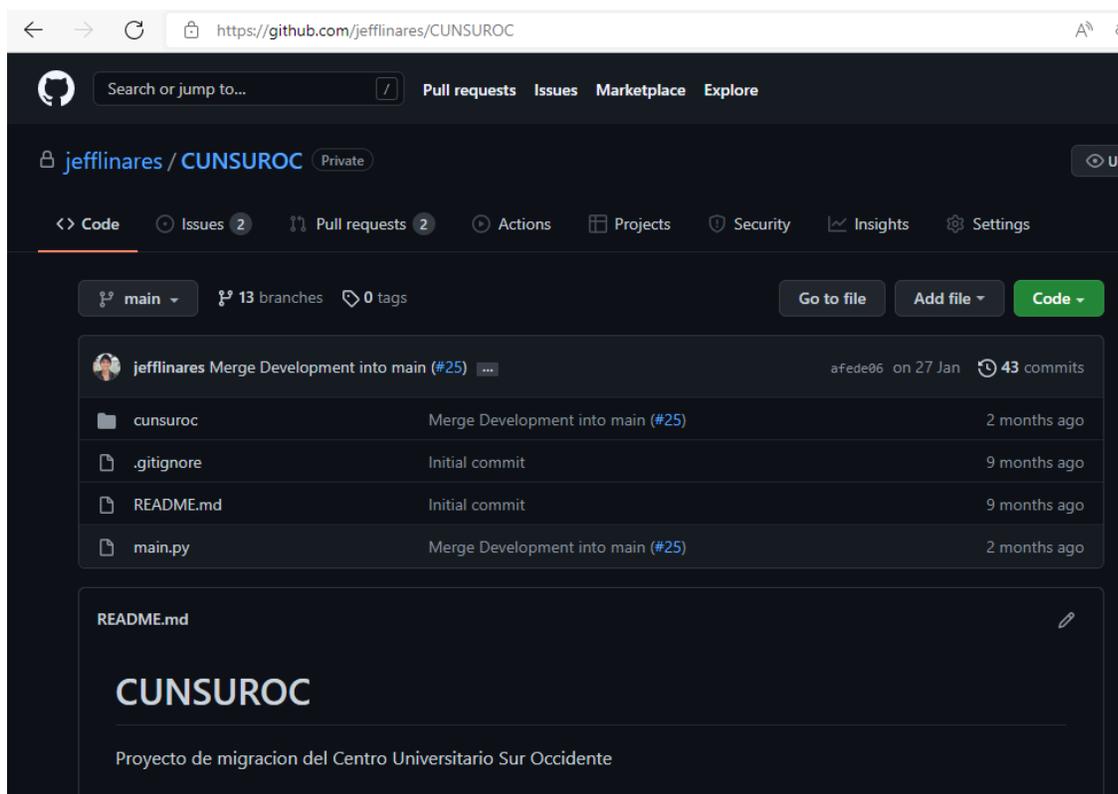
Todo el código se almacenó en un repositorio privado en la plataforma git, asociado a mi cuenta personal, en donde también se creó un proyecto privado para tener el control del mismo por medio de un tablero estilo Kanban en el que se manejaba la metodología scrum para llevar el control de las funcionalidades por hacer, las que se encontraban en progreso y las funcionalidades finalizadas, de esta forma se fue avanzando progresivamente en la integración de nuevas funcionalidades, corrección de bugs de una forma más ordenada y sencilla de interpretar.

En el directorio principal del repositorio se encontrará el archivo README.md en donde se explica la configuración necesaria para correr el

proyecto; inicia con la configuración e instalaciones necesarias para ejecutarse. También se define la estructura de este: es decir como se encuentra organizado por clases y librerías creadas, en donde se encuentran los archivos generados. Se define la estructura de algunos de los archivos utilizados para correr el mismo en relación con archivos \*.dbf como \*.csv, y la explicación de las funcionalidades más importantes utilizadas en el algoritmo ETL.

El acceso a dicho repositorio se otorgará a personal de CUNSUROC, y a personal de la Unidad de Procesamiento de Datos de la USAC.

Figura 9. Repositorio privado del CUNSUROC



Fuente: Github. *Repositorio proyecto CUNSUROC*. <https://github.com/jefflinares/CUNSUROC>.  
Consulta: marzo de 2022.

### **3.2.1. Manual técnico y documentación asociada**

Este manual contendrá la información detallada para establecer el ambiente local tanto para ejecutar el algoritmo del lenguaje de programación Python en el sistema operativo Windows 10, y la preparación de una base de datos local e instalación de un servidor para acceder a la misma a través del navegador; todo esto utilizando Docker como herramienta para el manejo de contenedores.



## CONCLUSIONES

1. Se definieron los archivos necesarios y la estructura que cada uno contenía para poder extraer la información y almacenarla en memoria en estructuras lógicas que representarían Clases en el lenguaje de programación Python.
2. El modelo de la base de datos destino se analizó y se plasmó en un diagrama entidad relación; con ello se logró tener una perspectiva visual de todas las tablas y campos necesarios para poder registrar de manera correcta los registros de cursos aprobados del CUNSUROC a la base de datos del nuevo Sistema de Control Académico.
3. El proceso de migración fue realizado elaborando un algoritmo programado en Python basado en la filosofía de un ETL; en él se implementó la fase de extracción: en la que se leyeron diferentes archivos: carreras, cursos, notas, entre otros, con el fin de relacionar la información; en la fase de transformación: ya con los datos en memoria, sufrieron las modificaciones necesarias para ajustarse a los campos requeridos por la nueva base de datos, y finalmente se generaron los archivos de inserción de todas las tablas involucradas con los registros consistentes.
4. Se realizó el proceso de migración de los archivos generados a un ambiente de pruebas de la base de datos, y en la página web del Sistema de Control Académico se realizaron diferentes auditorías con usuarios de pruebas proveídos por la Unidad de Procesamiento de Datos, y en dicho sistema se corroboró la información existente en CUNSUROC, validando

y corrigiendo los registros migrados y así realizar la migración a la base de datos de producción con la información consistente.

## RECOMENDACIONES

1. Planear una migración tomando en consideración lo siguiente, si las bases de datos son totalmente diferentes, se recomienda documentarse y estudiar ambas bases de datos; como, por ejemplo: estructura del lenguaje tanto para lectura y escritura de información, conocer el proceso de generar copias de seguridad, principalmente de la base de datos destino, dado a que sobre ella se cargaron los nuevos datos y de esta forma se podrá evitar que la nueva información genere algún conflicto con la información ya existente.
2. Analizar la tabla principal de la base de datos de destino y en base a ella definir las relaciones de sus llaves foráneas y de esta manera involucrar todas las tablas necesarias con el fin de que la información migrada sea consistente.
3. Realizar un proceso de ETL cuando los datos a migrar son demasiados y sobre todo utilizar herramientas de tecnología que permitan leer y escribir ambas bases de datos, en caso contrario es recomendable escoger un lenguaje de programación que brinde diferentes librerías que faciliten el proceso de codificar un ETL, en este caso Python me pareció bastante útil y práctico, ofreciendo librerías para leer archivos con extensión \*.dbf fácilmente, para la generación de los archivos de inserción representar cada tabla de la base de datos destino como una Clase de Python en la que se definan todos los campos necesarios para generar un registro válido, y almacenarlos en diccionarios generando una llave primaria para cada registro, con esto el acceso a los mismos será

mucho más rápido en lugar de utilizar listas o arreglos en donde se tendría que realizar una búsqueda secuencial si no se quisiera aplicar algún otro algoritmo de búsqueda avanzado.

4. Generar los archivos de inserción hacia la base de destino SQL, asegurándose que las sentencias de inserción de múltiples registros no superen los 2 mil registros, de lo contrario esto puede provocar algún error en la base de datos al sobrecargar dichas consultas, e inclusive ralentizar el proceso de inserción en el motor de la base de datos, al momento de ejecutar los procesos de migración realizarlo primero a una base de datos de prueba y una vez verificado iniciar el mismo proceso a la base de datos de producción, buscando un horario en donde la base de datos tenga poca o nula actividad. Para evitar que surjan bloqueos o errores inesperados, o que existan datos que se puedan perder al no haberse generado una copia de seguridad de estos, y sobre todo si la base de datos permite transaccionalidad, utilizarla y con ello se asegura que se migren todos los datos o ninguno en caso de existir algún error.

## BIBLIOGRAFÍA

1. DATE, C.J. *Introducción a las bases de datos*. 7a ed. Naucalpan de Juárez, México: Pearson Education, 2001. 960 p.
2. Docker Orientation and Setup. *Docker docs*. [en línea]. <<https://docs.docker.com/get-started/>>. [Consulta: septiembre de 2021].
3. Google Cloud Documentos. *Migración de bases de datos: conceptos y principios (parte 1)*. [en línea]. <<https://cloud.google.com/architecture/database-migration-concepts-principles-part-1?hl=es-419>>. [Consulta: agosto 2021].
4. LEGUIZAMON TARAZONA, Alba Yolima. *Pautas para una correcta migración de bases de datos*. [en línea]. <<https://repository.unilibre.edu.co/handle/10901/11161>>. [Consulta: agosto de 2021].
5. Milojičić, Dejan S. Douglis, Fred y Paindaveine, Yves. (2000). (Vol. 32). *Process migration*. NY USA. ACM Computing Surveys, 2000. pp. 241-299.
6. Python dbfread Introduction. *Read the docs*. [en línea]. <<https://dbfread.readthedocs.io/en/latest/introduction.html>>. [Consulta: octubre de 2021]

7. Universidad de San Carlos de Guatemala, Centro Universitario de SuroOccidente, CUNSUROC. *Acta estudiantil del curso de Matemática, referente al periodo de vacaciones del año de 1990.*  
002VAC90.dbf.

## APÉNDICES

Se explica brevemente la preparación y solución brindada a nivel técnico del algoritmo ETL para la realización de la migración de datos.

### Apéndice 1. **Ambiente de desarrollo local**

El ambiente de desarrollo local básicamente consta de 3 herramientas importantes, Instalación de Python y librerías utilizadas, IDE para el desarrollo y ejecución del código, Instalación de Docker con contenedores de base de datos MySQL.

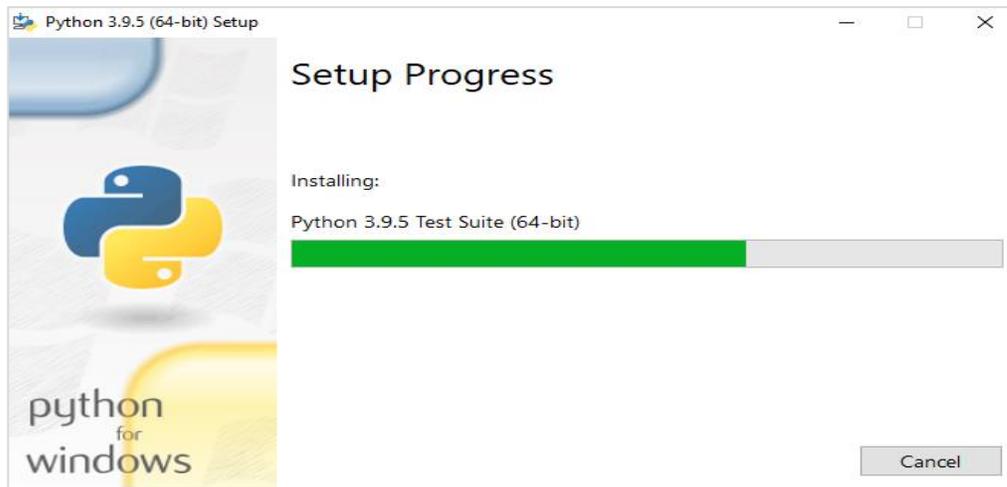
Para realizar la instalación y preparación del ambiente local en el sistema operativo Windows 10, se necesita una máquina con los siguientes requerimientos mínimos:

- Memoria RAM de 8gb
- Procesador i5 o i7 de 7ma generación (preferiblemente de 64 bits)
- Disco Duro con al menos 100 Gb disponibles

Instalación de Python: para ello se necesita ingresar al sitio web oficial <https://www.python.org/downloads/release/python-395/> para poder descargar la versión utilizada en el proyecto 3.9.5, se debe descargar el instalador dependiendo de los bits que soporta el procesador de la computadora, se debe ejecutar dicho instalador, y seguir los pasos que se desplegarán en pantalla.

Continuación del apéndice 1.

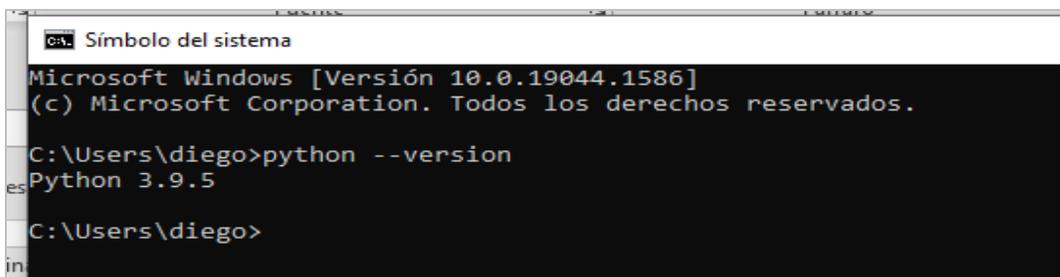
### Proceso de instalación de Python



Fuente: elaboración propia, empleando Instalador Python 2022.

Una vez instalado Python se puede comprobar la versión instalada del mismo, en la consola Símbolo del Sistema (CMD), del sistema operativo ejecutando el siguiente comando: `python --version` ó `py --version`

### Verificación de versión en CMD



Fuente: elaboración propia, empleando CMD.



## Continuación del apéndice 1.

Instalación de librería dbfread: para ello en la consola de Símbolo de Sistema de Windows se ejecuta el comando: `pip install dbfread`, para más información y acceder a la documentación de esta puede visitar la página oficial: <https://pypi.org/project/dbfread/>, esta librería permitirá leer los archivos de la base de datos del CUNSUROC, y con ello se puede extraer la información para generar el primer proceso del ETL, se habla del proceso de extracción, para poder continuar con la transformación y generación de los archivos con extensión \*.sql para finalmente realizar el proceso de carga de estos archivos a la base de datos de prueba y finalmente a la base de datos destino.

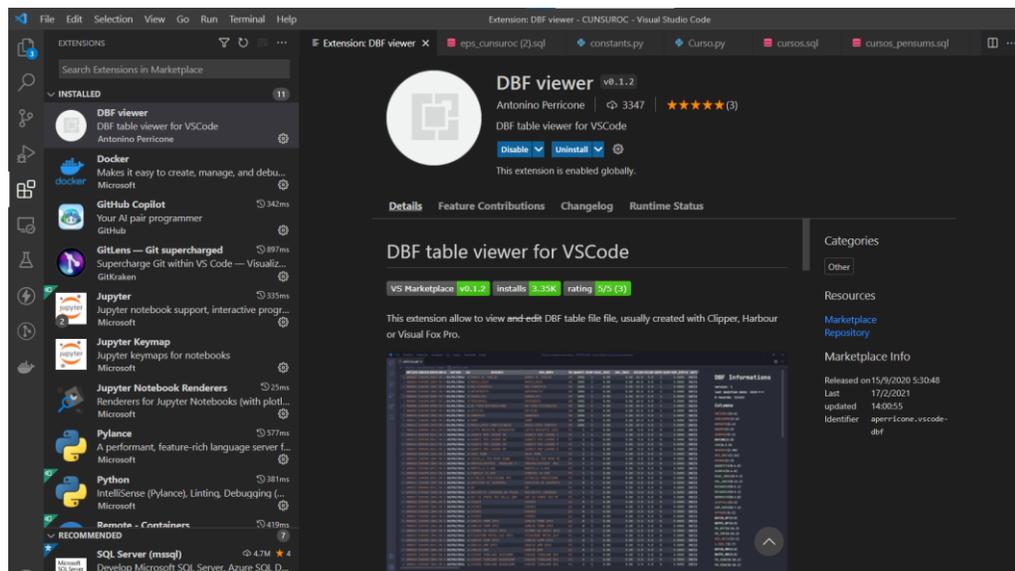
Instalación de IDE, en este caso se utilizó Visual Studio Code para el entorno de desarrollo, se deberá descargar desde el sitio oficial el instalador correcto para el sistema operativo, <https://code.visualstudio.com/download>.

Una vez instalado se deberá ejecutar e instalar algunas extensiones relacionadas con Python, estas son:

- *Python*: usada para ejecutar código Python en el IDE
- *Pylance*: permite formatear el código para hacerlo más legible y ordenado
- *DBF viewer*: usado para abrir archivos DBF de forma legible
- *GitLens*: permite visualizar en el código el autor y fecha de cuando se agregó dicho código en el repositorio relacionado a un commit.

Continuación del apéndice 1.

## Extensiones instaladas en Visual Studio Code



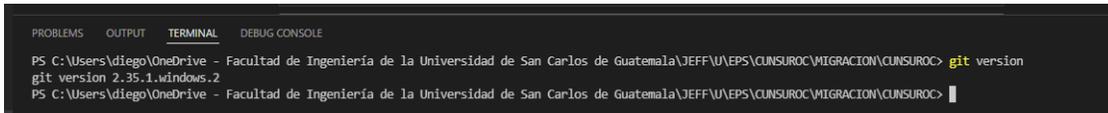
Fuente: elaboración propia, *Extensiones instaladas*, empleando Visual Studio Code 2022.

Instalación de Git: para manejo de un repositorio al cual se irán agregando las funcionalidades creadas, o correcciones al algoritmo ETL generado, se puede descargar desde la página oficial <https://git-scm.com/download/win>, se recomienda descargar el instalador en una versión no menor a la 2.32.0, al ejecutar el instalador se deberán dejar por defecto todas las configuraciones.

Reiniciar Visual Studio Code para que permita reconocer git en la terminal que este IDE ofrece, y así se puede verificar la versión, ejecutando el comando `git version`.

Continuación del apéndice 1.

## Verificación de versión de Git en Visual Studio Code



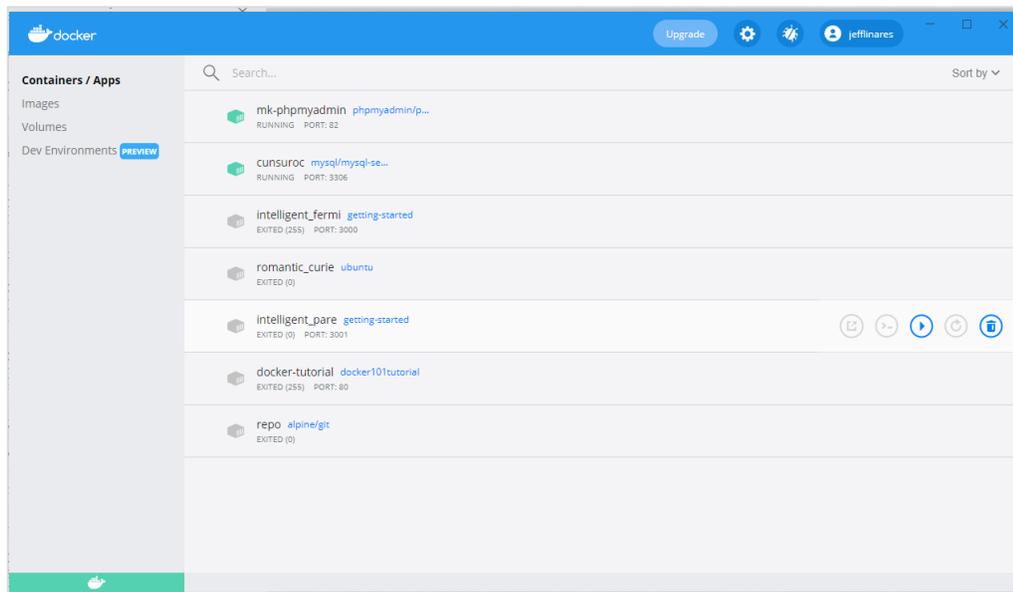
```
PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
PS C:\Users\diego\OneDrive - Facultad de Ingeniería de la Universidad de San Carlos de Guatemala\JEFF\U\EPS\CUNSUROC\MIGRACION\CUNSUROC> git version
git version 2.35.1.windows.2
PS C:\Users\diego\OneDrive - Facultad de Ingeniería de la Universidad de San Carlos de Guatemala\JEFF\U\EPS\CUNSUROC\MIGRACION\CUNSUROC> |
```

Fuente: elaboración propia, empleando Visual Studio Code console.

Instalación de Docker versión de escritorio: Docker servirá para crear contenedores de imágenes de MySQL y phpMyAdmin, en estos se generó una copia del modelo de Base de Datos del Sistema de Control Académico, esto con el fin de realizar pruebas en esta base de datos antes de afectar ambientes productivos de la base de datos de destino, se deberán seguir los pasos del sitio oficial de Docker para poder ejecutar correctamente este programa <https://docs.docker.com/desktop/windows/install/>.

Continuación del apéndice 1.

## Docker Desktop instalado

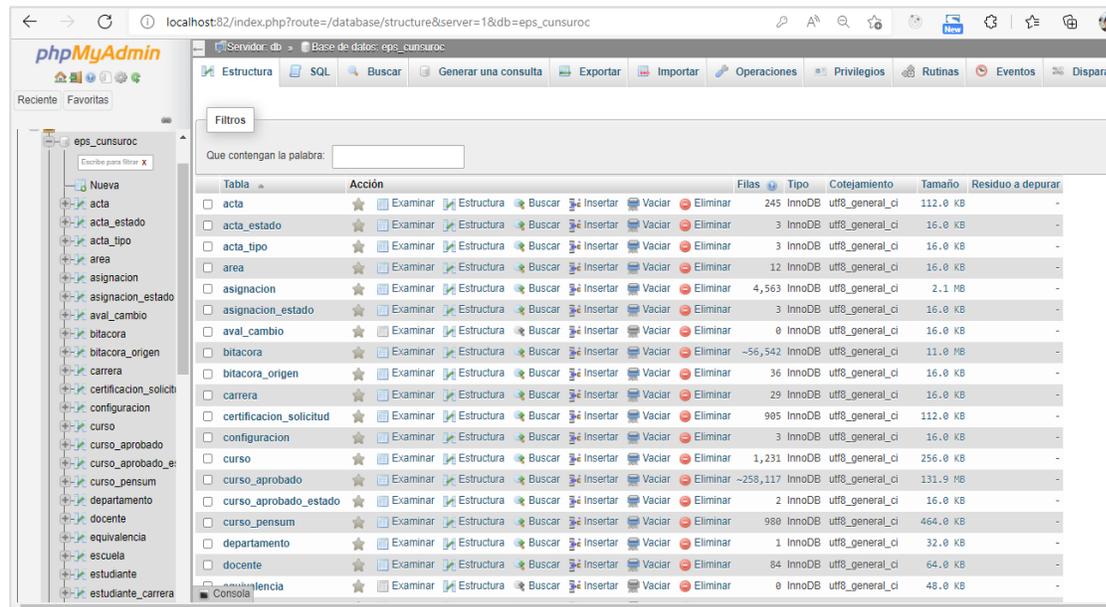


Fuente: elaboración propia, *Contenedores locales corriendo*, empleando Docker Desktop 2021.

Una vez instalado correctamente se deberá crear dos contenedores, un contenedor con un volumen definido para la base de datos MySQL y otro contenedor donde se tendrá configurado phpMyAdmin para tener acceso a dicha base de datos por medio del navegador, que por defecto correrá en el puerto 82, y la base de datos en el puerto 3306, sobre la creación de los contenedores y los comandos necesarios para realizarlo es necesario seguir los pasos definidos en el manual técnico que podrá encontrar en el siguiente enlace: [https://drive.google.com/drive/folders/1-9ldbGclsRR4C0-V\\_elq33qFHI-b7zK\\_?usp=sharing](https://drive.google.com/drive/folders/1-9ldbGclsRR4C0-V_elq33qFHI-b7zK_?usp=sharing).

Continuación del apéndice 1.

## phpMyAdmin ejecutándose de forma local en el puerto 82

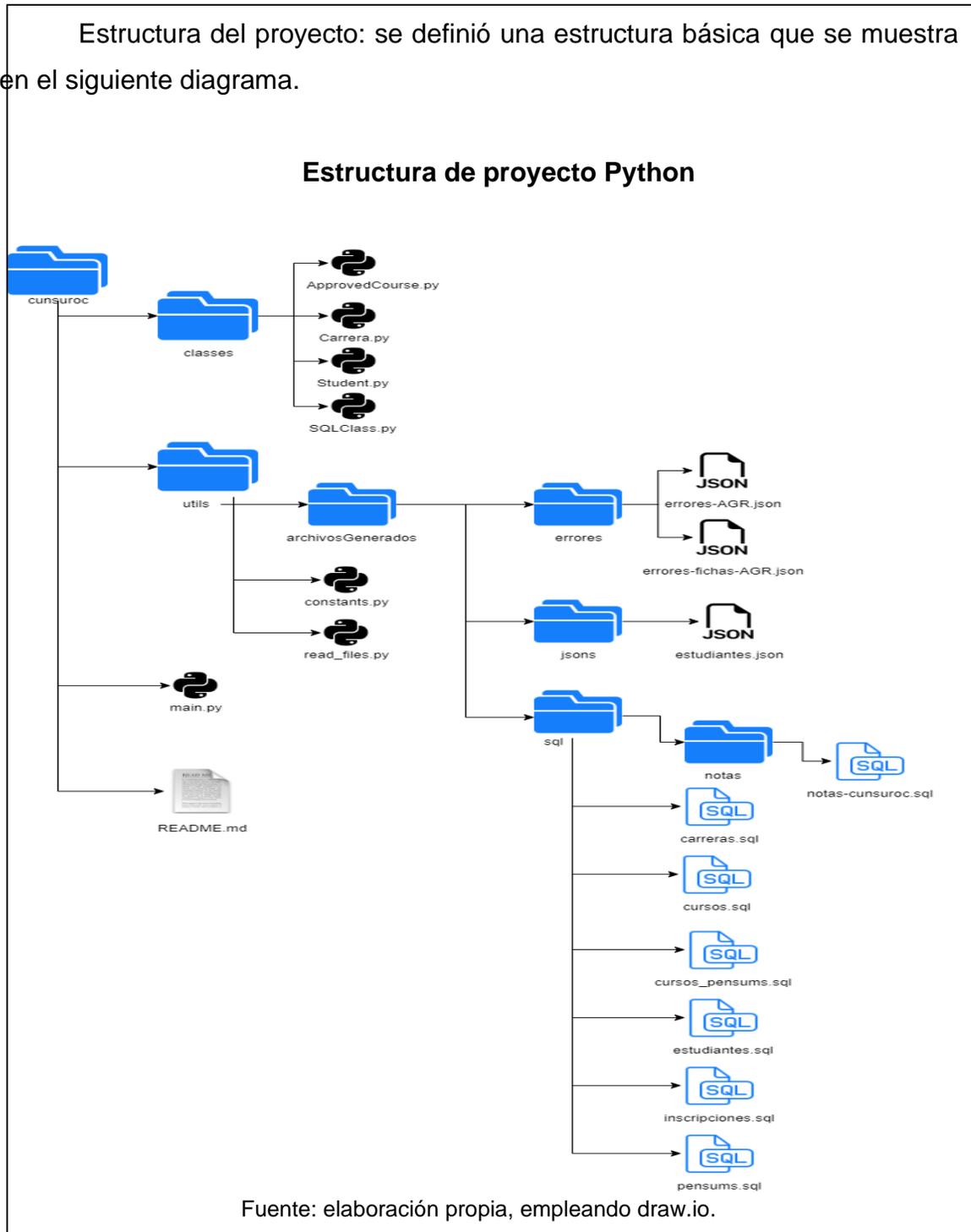


Fuente: elaboración propia, empleado Microsoft Edge.

Y con ello se finaliza la configuración del ambiente de desarrollo local necesario para iniciar el algoritmo ETL en Python.

## Apéndice 2. Algoritmo ETL en Python

Estructura del proyecto: se definió una estructura básica que se muestra en el siguiente diagrama.



## Continuación del apéndice 2.

El proyecto cuenta con una estructura básica en la que en la raíz de este se encuentra la función principal que ejecuta la secuencia correcta para iniciar el proceso ETL, la carpeta classes contiene la definición de todas las clases utilizadas en el proyecto en las que destaca una clase principal llamada SQLClass que permite generar la salida de sus atributos en formato json o en una estructura sql, esta clase se utiliza como clase principal de la que heredan otras clases utilizadas que se listan a continuación:

- **ApprovedCourse:** contiene atributos que permite registrar datos en la tabla curso\_aprobado como: carnet, carrera, pensum, curso, entre otros.
- **Carrera:** contiene atributos necesarios para realizar la inserción a la tabla carrera como: carrera, nombre, unidad, extensión, entre otros.
- **Curso:** esta clase permite guardar campos relacionados a la tabla curso como: código, nombre, carrera, pensum, entre otros.
- **Error:** esta clase permitirá registrar cualquier tipo de error encontrado ya sea técnico o semántico en el proceso ETL, como por ejemplo registrar errores de archivos que no cuenten con campos obligatorios como lo son carnet del estudiante, o que en los valores de notas contengan un valor no permitido.
- **Student:** esta clase contiene atributos que permiten registrar la información relacionada con los estudiantes como: carnet, cui, nombre, sexo, passwd, entre otros, que posteriormente se registrará en la tabla estudiante.

## Continuación del apéndice 2.

También se crearon otras clases que no heredan de SQLClass, pero son igual de importantes como el resto, entre ellas destacan:

- Record: esta clase permite formatear cada fila de los archivos \*.dbf estableciendo por cada columna un objeto con atributos nombre y valor y también permitirá manejar los registros que sean inválidos y no se logren leer correctamente.
- Pensum: permite registrar información relacionada a los pensum como: pensum, carrera, vigencia\_inicio, creditos\_cierre
- utils.py: este archivo contendrá diferentes funciones que permitirán transformar la información de origen, como por ejemplo comparar cadenas, validar si un valor es del tipo numérico, entre otras.

Dentro de la carpeta utils se encuentra la carpeta archivosGenerados que en ella se encontrarán todos los archivos de inserción \*.sql, también los errores encontrados por cada carpeta y fichas leídas, los errores generales del proceso ETL, y de igual forma se encontrarán los siguientes archivos:

- constants.py: en este archivo se define las rutas específicas donde se tendrá que ir a leer la base de datos de origen y algunos archivos principales como, por ejemplo: el archivo donde se encuentra la información de las carreras con extensión \*.csv; el archivo donde se definen los cursos y pensum asociados a este centro regional; la ruta de la carpeta donde se encuentran las notas de todas las carreras; el nombre de la carpeta que contiene las fichas; se configura también las rutas donde

## Continuación del apéndice 2.

se generaran los archivos de salida \*.sql, los archivos \*.json; entre otras constantes definidas siguiendo las reglas del negocio.

- read\_files.py: este es el archivo que contiene los diferentes procesos traducidos a funciones de algoritmos de programación para lograr la ejecución del ETL, cuenta con 3 secciones principales:
  - Sección de importaciones: se importan todas las clases, constantes, y funciones de otros archivos que se utilizarán, se importa también la librería dbfread para poder leer cada uno de los archivos \*.dbf.
  - Sección de definición de variables: se definen todas las variables globales, la definición de las carpetas de cada carrera que deberán ser leídas especificando el archivo principal de cada una; dicho archivo contiene la información de las actas que deben ser leídas dentro de la misma.
  - Sección de definición de funciones: entre las principales destacan las siguientes:
    - writeFile: función utilizada para escribir un archivo en donde se envía el nombre del archivo, el contenido de este, y la ruta donde será escrito como parámetros.

## Continuación del apéndice 2.

- readCareers: función que lee el archivo de las carreras.csv y crea su objeto correspondiente para formatear la información.
- readPensum: esta función permite leer el archivo donde están definidos la mayoría de los cursos y pensums impartidos en todas las carreras del CUNSUROC, para poder mantener en memoria estos registros que posteriormente serán utilizados.
- readStudentsCSV: función que permite comparar la información de estudiantes de Registro y Estadística de un archivo \*.json contra la información de estudiantes del CUNSUROC y con ello poder obtener más datos como correo, teléfono, sexo, cui, entre otros campos que no son proveídos en la base de datos origen.
- readFolders: esta función sin duda alguna es la más importante de todas, en ella se realiza el proceso de lectura de las carpetas que contiene los cursos aprobados por carreras, el algoritmo lee cada carpeta, busca su archivo principal lo lee y lo recorreré este archivo contiene el nombre de las actas que contiene la información de los cursos aprobados, así que se lee cada registro y posteriormente se busca este archivo y se lee extrayendo las notas de la misma, recorriendo registro por registro, registrando información de estudiantes, cursos si no existiera definido previamente, y

## Continuación del apéndice 2.

registros de cursos aprobados, al finalizar de recorrer cada carpeta ejecuta un método para escribir los archivos de inserción \*.sql referente a las notas de esta carrera, y los errores que se encontraron en la misma. Una vez finalizado todo el proceso de lectura de todas las carreras, se ejecutan procesos de escritura de los demás archivos de inserción referentes a carreras, cursos, pensum, estudiantes, inscripciones, entre otros.

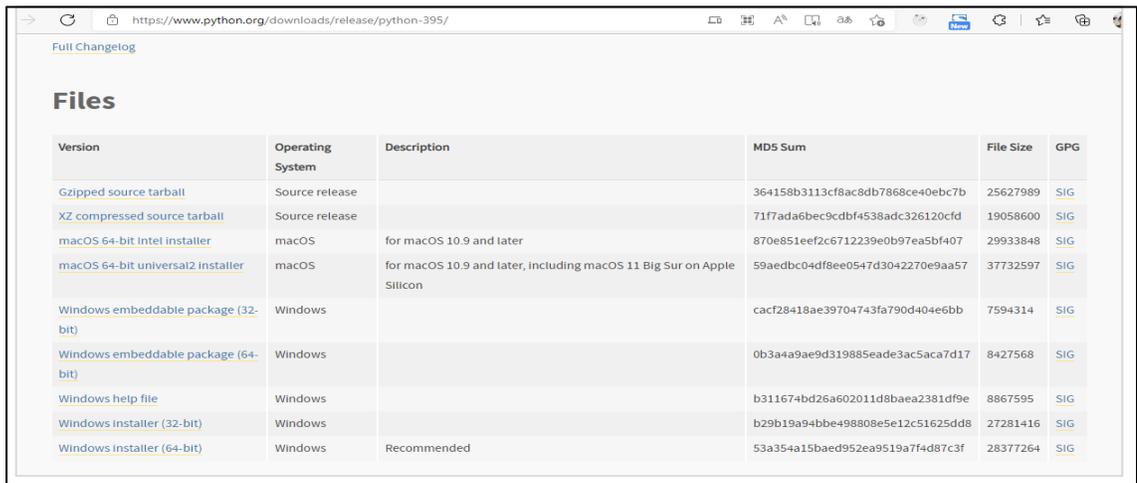
- registerInscription: función que registra la inscripción de un estudiante a una carrera en específico, formateando y almacenando la data necesaria para generar el archivo de salida perteneciente a inscripciones.

Fuente: elaboración propia.

## ANEXO

A continuación, se presentan algunas imágenes de la instalación y configuración del ambiente local para llevar a cabo la migración.

### Anexo 1. Descarga de Python 3.9.5 del sitio oficial



The screenshot shows a web browser window with the URL <https://www.python.org/downloads/release/python-395/>. The page title is "Full Changelog" and the main heading is "Files". Below the heading is a table with the following data:

Version	Operating System	Description	MD5 Sum	File Size	GPG
<a href="#">Gzipped source tarball</a>	Source release		364158b3113cf8ac8db7868ce40ebc7b	25627989	<a href="#">SIG</a>
<a href="#">XZ compressed source tarball</a>	Source release		71f7ada6bec9c9dbf4538adc326120cfd	19058600	<a href="#">SIG</a>
<a href="#">macOS 64-bit Intel installer</a>	macOS	for macOS 10.9 and later	870e851eef2c6712239e0b97ea5bf407	29933848	<a href="#">SIG</a>
<a href="#">macOS 64-bit universal2 installer</a>	macOS	for macOS 10.9 and later, including macOS 11 Big Sur on Apple Silicon	59aedbc04df8ee0547d3042270e9aa57	37732597	<a href="#">SIG</a>
<a href="#">Windows embeddable package (32-bit)</a>	Windows		cacf28418ae39704743fa790d404e6bb	7594314	<a href="#">SIG</a>
<a href="#">Windows embeddable package (64-bit)</a>	Windows		0b3a4a9ae9d319885eade3ac5aca7d17	8427568	<a href="#">SIG</a>
<a href="#">Windows help file</a>	Windows		b311674bd26a602011d8baea2381df9e	8867595	<a href="#">SIG</a>
<a href="#">Windows installer (32-bit)</a>	Windows		b29b19a94bbe498808e5e12c51625dd8	27281416	<a href="#">SIG</a>
<a href="#">Windows installer (64-bit)</a>	Windows	Recommended	53a354a15baed952ea9519a7f4d87c3f	28377264	<a href="#">SIG</a>

Fuente: Python. *Descarga de Python*. <https://www.python.org/downloads/release/python-395/>.

Consulta: marzo de 2022.

