



## PROYECTO NO. 2

### 1. OBJETIVOS

- **General**

- Utilizar consultas para presentar la información requerida de una base de datos relacional.

- **Específicos**

- Desarrollar la habilidad de generar consultas avanzadas en lenguaje SQL.
- Mostrar información específica de una base de datos.
- Aprender a realizar cargas masivas desde archivos separados por coma (csv) a una base de datos.
- Desarrollar el modelo entidad relación a partir de un conjunto de datos.

### 2. DESCRIPCIÓN DEL PROYECTO

#### 2.1 ENUNCIADO 1

El centro de datos lo contrata a usted para realizar reingeniería al proceso de almacenamiento que actualmente manejan para guardar la información de las diferentes compañías que tiene a cargo ya que actualmente hay demasiada inconsistencia que puede provocarles serios problemas legales. El centro de datos hace el esfuerzo de exportar toda la información actual a un archivo delimitado por comas (csv) que usted tendrá a disposición para realizar el análisis correspondiente con el objetivo de eliminar redundancia de los datos y crear un nuevo modelo de base de datos estable para almacenar la información.

#### ESTRUCTURA DEL ARCHIVO CSV

A continuación, se muestra cómo está estructurado el archivo separado por comas que contiene los datos a cargar. Con esta información se debe crear el modelo de datos que permita realizar los reportes. El archivo puede ser descargado desde el UEDI en la sección Proyectos.

Columna	Descripción
nombre_compania	Nombre de la compañía <i>Tipo: texto</i>
contacto_compania	Nombre de la persona con la que se tiene contacto dentro de la empresa. <i>Tipo: texto</i>

correo_compania	Correo electrónico de la compañía. <i>Tipo: texto</i>
telefono_compania	Teléfono de la compañía. <i>Tipo: texto</i>
tipo	Carácter que representa si los siguientes campos (nombre, correo, teléfono y campos de dirección) corresponde a un Cliente [C] o a un proveedor [P]. <i>Tipo: carácter</i>
nombre	Nombre del cliente ó proveedor. <i>Tipo: texto</i>
correo	Correo electrónico del cliente ó proveedor. <i>Tipo: texto</i>
teléfono	Teléfono del cliente ó proveedor. <i>Tipo: texto</i>
fecha_registro	Fecha en la que se registro al sistema el cliente ó proveedor. <i>Tipo: fecha</i>
dirección	Dirección del cliente ó proveedor. <i>Tipo: texto</i>
ciudad	Ciudad de donde es el cliente ó proveedor. <i>Tipo: texto</i>
codigo_postal	Código postal de la dirección del cliente ó proveedor. <i>Tipo: numérico</i>
región	Región correspondiente de la ciudad de donde es el cliente ó proveedor. <i>Tipo: texto</i>
producto	Nombre del producto que se compró/vendió según corresponda. <i>Tipo: texto</i>
categoria_producto	Categoría a la que pertenece el producto. <i>Tipo: texto</i>

cantidad	Número que representa la cantidad de producto adquirido/vendido. <i>Tipo: numérico</i>
precio_unitario	Costo que representa el adquirir una única unidad de un producto en específico. <i>Tipo: decimal</i>

## SOLICITUD

Una vez analizado el archivo anterior, se solicita que usted diseñe un nuevo modelo de datos para almacenar toda esa información. El proceso es el siguiente:

1. Debe crear una tabla temporal en donde debe cargar toda la información del CSV.
2. Haciendo uso de consultas a la tabla temporal debe de llenar el modelo que usted propone para almacenar la información de forma correcta de modo que no exista redundancia de los datos.
3. Una vez lleno todo el modelo, debe realizar los reportes siguientes los cuales serán evaluados para verificar que la información mostrada sea consistente y verídica.

## REPORTES

Se deben de entregar los siguientes reportes:

1. Mostrar el número de cliente, nombre, apellido y total del cliente que más productos ha comprado.
2. Mostrar el número de mes de la fecha de registro, nombre y apellido de todos los clientes que más han comprado y los que menos han comprado (en dinero) utilizando una sola consulta.
3. Mostrar el top 5 de proveedores que más productos han vendido (en dinero) de la categoría de productos **‘Fresh Vegetables’**.
4. Mostrar el nombre del proveedor, número de teléfono, número de orden, total de la orden por la cual se haya obtenido la menor cantidad de producto.
5. Mostrar el top 10 de los clientes que más productos han comprado de la categoría **‘Seafood’**.
6. Mostrar el porcentaje de clientes que le corresponden a cada región.
7. Mostrar las ciudades en donde más se consume el producto **“Tortillas”** de la categoría **“Refrigerated Items”**.
8. Mostrar la cantidad de clientes que hay en las ciudades agrupadas por su letra inicial, es decir, agrupar las ciudades con A y mostrar la cantidad de clientes, lo mismo para las que inicien con B y así sucesivamente.
9. Mostrar el porcentaje de las categorías más vendidas de cada ciudad de la siguiente manera: Ciudad, Categoría, Porcentaje De Mercado
10. Mostrar los clientes que hayan consumido más que el promedio que consume la ciudad de Frankfort.

## 2.2ENUNCIADO 2

La comisión de patentes mundial CPM es una organización que aglutina a muchos países del mundo y cuya función es llevar un control de las patentes y los inventos que se realizan en sus países miembros. La CPM ha manejado esta información a través de una hoja de cálculo (Excel) por lo que ahora se quiere que realice un sistema que sustituya la hoja de cálculo y elimine la redundancia de información.

Cada invento que se realiza es asociado a un país. Cuando se patenta el invento este es asociado a cada país. Si un invento es inventado por más de un inventor de igual manera el invento es asociado a un solo país. Los inventores deben indicar a cual en la patente, tradicionalmente se asigna al país donde se realizó el invento, independiente de la nacionalidad del inventor pero no es una regla, De igual manera, es importante la nacionalidad de cada inventor, Para la CPM es importante el área de cada país y su población pues un factor importante de desarrollo es saber cuántos inventos se realizan por área, y cuantos por

habitante. Muchas veces la influencia de un invento no solo se da en el país donde se realizó el invento sino también en los países cercanos por lo que para la CPM es importante conocer las fronteras de cada país. Los países están ubicados en regiones. Las regiones pueden estar contenidas en otras regiones. Así por ejemplo, Guatemala está en Centro América, que a su vez está en América. Las regiones son las siguientes:

La CPM está conformada por un equipo de profesionales de diferentes áreas de investigación científica, como por ejemplo, Óptica, Mecánica, Energía, Electrónica, Biología, Medicina, Química, Matemática etc. Cada área de investigación tiene una descripción muy específica con la que los profesionales deben trabajar, así como un "ranking" asignado por los profesionales. El ranking es un número entero entre 0 y 10 que define el grado de tiempo de trabajo que requiere. Así, si un área tiene 10 significa que consume mucho tiempo de trabajo en comparación con una que tiene 1. Muchos profesionales dominan varias áreas de investigación sin ningún problema. Cada profesional es asignado a documentar y estudiar cada invento, un profesional puede trabajar en varios inventos al mismo tiempo. Todos los profesionales tienen un salario y algunos pueden tener una comisión asignada por la CPM. Es importante conocer la fecha de inicio de funciones en la CPM pues la experiencia es importante en este tipo de trabajos. Cada área de investigación tiene un profesional como jefe y además hay un profesional como Jefe de todas las áreas. Cada país miembro tiene un representante en la CPM. La CPM realiza periódicamente una serie de encuestas a los países miembros para conocer el punto de vista de los países sobre los inventos. Las encuestas pueden ser de diferentes temas asociados, por ejemplo, Sobre inventos, sobre los profesionales de la CPM, el desarrollo de los países, sobre historia de los inventos, etc. Cada encuesta está formada por una serie de preguntas de selección múltiple. Cada pregunta tiene una serie de posibles respuestas. Cada país responde las preguntas en base a la respuesta que considere más adecuada.

## REQUERIMIENTOS

### Documentación

La CPM solicita que realice un documento de análisis que le permita determinar el alcance de la nueva plataforma. El documento deberá tener como mínimo una descripción de las reglas de normalización aplicadas al archivo de Excel fuente, modelo entidad relación explicativo que permita visualizar de mejor manera la nueva estructura y de cómo se almacenará la información.

### Scripts

La CPM solicita que se entregue el script completo de la base de datos para la nueva plataforma, creación de tablas, constraint, llaves primarias, candidatas únicas, check de referencias que usted considere necesario.

### Carga de datos

La CPM hace entrega de un archivo Excel con toda la información de su sistema desnormalizada. Se solicita que se genere un script que permita cargar la información del archivo de exportación (Excel) a una tabla temporal y con base al modelo entidad relación solicitado en la documentación y ejecutado mediante el script, distribuya toda la información hacia la nueva estructura propuesta. Los archivos de carga lo encontrarán en el apartado Proyecto 2 en la carpeta Enunciado 2.

### Reportes

La CPM tiene diez (10) consultas de uso frecuente. Se adjuntara el requerimiento de las diez consultas para que las pueda implementar con base a la nueva estructura propuesta. Deberá crear un script con las diez consultas.

## REQUERIMIENTOS Y RESTRICCIONES

- Deben utilizar DataModeler para diseñar el modelo relacional.
- El proyecto debe de implementarse en MySQL.
- No se debe separar en diferentes archivos csv el archivo proporcionado. Debe ser cargado únicamente el que fue proporcionado.
- Los reportes deben ser realizados a la entidad relación creada por el estudiante, reportes hechos a la tabla temporal serán anulados.
- El proyecto es individual.
- Copias totales o parciales tendrán nota de 0 puntos y serán reportados a escuela.

## ENTREGABLES

### Enunciado 1

- El modelo relacional en notación baker propuesto en formato de imagen. Formato: **[BD1]ModeloPropuesto.png**
- Script (con extensión .sql) que contiene las instrucciones DDL para crear el modelo relacional. A este script debe de aplicarle todas las reglas de integridad vistas en clase y laboratorio. Formato: **[BD1]InstruccionesDDL.sql**
- Script (con extensión .sql) que contiene todas las consultas necesarias para llenar el modelo relacional propuesto a partir de la tabla temporal. Formato: **[BD1]CargaDeDatos.sql**
- Script (con extensión .sql) que contiene todas las consultas necesarias para realizar los reportes. Formato: **[BD1]Consultas.sql**

### Enunciado 2

- Documentación de análisis solicitada
- Script (ddl) para la creación de la base de datos
- Script para la carga de datos y distribución de información en la nueva base de datos
- Script con las veinte consultas

Todos los archivos anteriormente mencionados deben de ir dentro de una carpeta comprimida con el siguiente formato:

**[BD1]Proyecto2\_#carnet.rar**

## FECHA DE ENTREGA

- Domingo 3 de enero de 2020, antes de las 23:59 am.
- La entrega será a través de la plataforma de UEDI en el post correspondiente al Proyecto No. 2.
- Los horarios de calificación serán notificados en el transcurso de las semanas siendo el lunes 4 la calificación del mismo.