

Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Base de Datos 1
Ing. Luis Espino / Aux. Gerardo Chay
Segundo Semestre 2021



PROYECTO NO. 2

INTRODUCCIÓN

La necesidad de tener sistemas escalables y óptimos depende principalmente de un buen diseño de base de datos. Si bien existen sistemas antiguos que funcionan así, con una metodología de almacenamiento de información en archivos planos con datos redundantes y con datos no atómicos, éstos a lo largo del tiempo llegan a impactar de forma considerable el rendimiento del propio sistema. Por lo anterior muchas organizaciones aceptan el reto de migrar sus sistemas a bases de datos formalizadas que les permita mejorar su rendimiento y su escalabilidad.

OBJETIVOS

- Diseñar y desarrollar un modelo relacional a partir del planteamiento de un problema y el análisis de un archivo de datos.
- Realizar una carga masiva de datos de la información proporcionada a una nueva plataforma mediante la creación de una tabla temporal distribuyendo la información de la carga al modelo relacional propuesto.
- Generar consultas avanzadas en lenguaje SQL que cumplan con los reportes solicitados.

ENUNCIADO

a. Preliminares

El Instituto Centroamericano Electoral es una institución dedicada a registrar, controlar y evaluar estadísticas de los comicios electorales en los diferentes países de Centro América, para lo cual requiere un sistema de bases de datos donde se puedan hacer consultas de diferentes temas electorales.

Los países están divididos en regiones. Cada región está formada por un conjunto de departamentos o provincias, y cada provincia tiene un conjunto de municipios. A cualquiera de estos (país, municipio, departamento o región) se le llama una zona. Para la institución no es importante llevar información de los datos de los ciudadanos, pues el voto es secreto. Sin embargo, es importante tener información sobre las características generales de la población para tomar estadísticas respecto al voto. Así, a la población se le puede clasificar de diferentes maneras, dependiendo del tipo de información que la institución quiera saber. Por ejemplo, por sexo: hombres, mujeres; por educación mínima: analfabetos, alfabetos; por raza: indígenas, ladinos, garífunas, etc.; por escolaridad: primaria, nivel medio, universitario; por edad: joven, adulto, tercera edad. Estos son solo ejemplos, pero la institución puede dividir a los votantes de la forma que considere adecuada para manejar información y tomar decisiones. De esta manera se puede saber si los jóvenes, o las mujeres o los analfabetos votan más, en qué país, municipio, departamento, etc. hay más votantes universitarios. Lo interesante además, es que se quiere llevar información de elecciones de diferentes años para hacer comparaciones. En cada elección es importante el año y el tipo de elección o el nombre que se le coloca en cada país. Por ejemplo elecciones generales, municipales, etc. del año 2007 en Guatemala.

Cada elección tiene un conjunto de puestos de elección popular que se definen en cada país y que abarcan una zona preestablecida (país, región, departamento o municipio). Por ejemplo, un puesto de elección en Guatemala puede ser de alcalde y por municipio, es decir, se eligen alcaldes para cada municipio. Otro puesto de elección es el de presidente, pero éste es por país. La elección de diputados es regional. La elección de gobernadores es departamental. Una elección puede tener elecciones de diputados, presidentes, alcaldes, gobernadores, etc. De tal forma que se vota por presidente en todo el país, pero para alcalde en cada municipio, así los ciudadanos que votan por un alcalde en su municipio no pueden votar por alcalde en otro municipio, por ejemplo.

Los ciudadanos votan para un puesto de elección por candidatos que deben, por ley, ser propuestos por partidos políticos o comités cívicos. De esta forma, los partidos políticos participan para ser electos en cualquier puesto de elección que quieran. Por ejemplo, el partido ABC participa en Guatemala, para elecciones de presidente y diputados, otros partidos participarán en otros puestos de elección. Esto es igual en todos los países. No es importante saber el nombre del candidato, sino solamente del partido político en cada país, que participa en una elección específica por un puesto de elección en una zona del país. Bajo este esquema, se puede saber cuántos votos obtuvo un partido político en determinada elección para determinado puesto de elección, en una zona dada y las características de los votantes (raza, escolaridad, sexo, etc.).

b. Solicitud

El Instituto Centroamericano Electoral (ICE) contrata sus servicios profesionales para realizar reingeniería al proceso de almacenamiento debido a que actualmente hay demasiada inconsistencia que puede provocarles serios problemas legales. El Instituto realiza el esfuerzo de exportar toda la información actual a un archivo Excel que usted tendrá a disposición para realizar el análisis correspondiente, deberá aplicar las formas de normalización que considere adecuada con el objetivo de eliminar la redundancia y crear una nueva plataforma escalable.

c. Requerimientos

i. Documentación

El ICE 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.

ii. Scripts

El ICE 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.

iii. Carga de datos

El ICE 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 el inciso **c.i** para luego ser ejecutado mediante el inciso **c.ii**. distribuyendo así toda la información hacia la nueva estructura propuesta. El archivo de carga lo encontrarán en el siguiente link: <https://tinyurl.com/yykzknzh>

iv. Reportes

1. Desplegar para cada elección el país y el partido político que obtuvo mayor porcentaje de votos en su país. Debe desplegar el nombre de la elección, el año de la elección, el país, el nombre del partido político y el porcentaje que obtuvo de votos en su país.
2. Desplegar total de votos y porcentaje de votos de mujeres por departamento y país. El ciento por ciento es el total de votos de mujeres por país. (Tip: Todos los porcentajes por departamento de un país deben sumar el 100%)

3. Desplegar el nombre del país, nombre del partido político y número de alcaldías de los partidos políticos que ganaron más alcaldías por país.
4. Desplegar todas las regiones por país en las que predomina la raza indígena. Es decir, hay más votos que las otras razas.
5. Desplegar el nombre del país, el departamento, el municipio y la cantidad de votos universitarios de todos aquellos municipios en donde la cantidad de votos de universitarios sea mayor que el 25% de votos de primaria y menor que el 30% de votos de nivel medio. Ordene sus resultados de mayor a menor.
6. Desplegar el porcentaje de mujeres universitarias y hombres universitarios que votaron por departamento, donde las mujeres universitarias que votaron fueron más que los hombres universitarios que votaron.
7. Desplegar el nombre del país, la región y el promedio de votos por departamento. Por ejemplo: si la región tiene tres departamentos, se debe sumar todos los votos de la región y dividirlo dentro de tres (número de departamentos de la región).
8. Desplegar el total de votos de cada nivel de escolaridad (primario, medio, universitario) por país, sin importar raza o sexo.
9. Desplegar el nombre del país y el porcentaje de votos por raza.
10. Desplegar el nombre del país en el cual las elecciones han sido más peleadas. Para determinar esto se debe calcular la diferencia de porcentajes de votos entre el partido que obtuvo más votos y el partido que obtuvo menos votos.
11. Desplegar el total de votos y el porcentaje de votos emitidos por mujeres indígenas alfabetas.
12. Desplegar el nombre del país, el porcentaje de votos de ese país en el que han votado mayor porcentaje de analfabetas. (tip: solo desplegar un nombre de país, el de mayor porcentaje).
13. Desplegar la lista de departamentos de Guatemala y número de votos obtenidos, para los departamentos que obtuvieron más votos que el departamento de Guatemala.

ENTREGABLES

- Documentación de análisis solicitada, con los pasos que siguió para realizar la normalización. Debe realizarse en el **README.md** global del repositorio.
- Script (DDL) para la creación de la base de datos. El formato es: **ScriptCreacion.sql**

- Script para la carga de datos y distribución de información en la nueva base de datos. El formato es: **CargaDatos.sql**
- Script con las consultas requeridas. El formato es: **Consultas.sql**

RESTRICCIONES

- El proyecto es individual.
- El sistema operativo queda a discreción del estudiante.
- El proyecto debe ser implementada en una base de datos **MySQL**.
- Las copias de prácticas tendrán automáticamente nota de 0 puntos y se reportará a los involucrados a la Escuela de Ingeniería en Ciencias y Sistemas.
- No se recibirán proyectos después de la fecha de entrega.
- Deben tener su proyecto en un repositorio **GitHub**
- El nombre del repositorio debe ser **bd1-proyecto2-#carnet**
- Deben agregar al auxiliar al repositorio, mi usuario es: **gerchay**

ENTREGA

- La entrega será por medio de la plataforma de EUDI.
- Tomar en cuenta que no habrá prórroga, por lo que se solicita al estudiante distribuir adecuadamente su tiempo para la elaboración de su práctica.
- Para entregar el proyecto deberán copiar el link de su repositorio donde tienen su proyecto 2.
- Fecha de entrega:

31 de octubre de 2021, antes de las 23:59