

Carrera: INGENIERIA EN INFORMATICA

Asignatura: 3641 – Bases de Datos Aplicada.

Tema: Bases de datos relacionales

Unidad: 2

Objetivo: Que el alumno practique algunos conceptos básicos de SQL que deberá manejar en el análisis de datos.

Competencia/s a desarrollar:

- Especificación, proyecto y desarrollo de sistemas de información.
- Especificación, proyecto y desarrollo de sistemas de comunicación de datos.
- Especificación, proyecto y desarrollo de software.
- Proyecto y dirección en lo referido a seguridad informática.
- Establecimiento de métricas y normas de calidad de software.
- Procedimientos y certificaciones del funcionamiento, condición de uso o estado de sistemas de información, sistemas de comunicación de datos, software, seguridad informática y calidad de software.
- Dirección y control de la implementación, operación y mantenimiento de sistemas de información, sistemas de comunicación de datos, software, seguridad informática y calidad de software.
- Identificación, formulación y resolución de problemas de ingeniería en sistemas de información/informática.
- Concepción, diseño y desarrollo de proyectos de ingeniería en sistemas de información / informática.
- Gestión, planificación, ejecución y control de proyectos de ingeniería en sistemas de información / informática.
- Utilización de técnicas y herramientas de aplicación en la ingeniería en sistemas de información / informática.
- Generación de desarrollos tecnológicos y/o innovaciones tecnológicas.
- Desempeño en equipos de trabajo.
- Comunicación efectiva.

Descripción de la actividad:

1. Tiempo estimado de resolución: 1 semana
2. Metodología: En computadora.
3. Forma de entrega: No obligatoria
4. Metodología de corrección y feedback al alumno: Presencial y por Miel.

Pauta general: Todos los ejercicios deben resolverse con DML/DDL. Aunque puede usar la GUI en caso de no recordar la sintaxis, recomendamos enfáticamente practicar y aprender la sintaxis de SQL.

Los ejercicios se pensaron para ser resueltos en Microsoft SQL Server. Puede usar la versión Developer o Express. Recomendamos emplear versión 2019 o 2022. Sin embargo, si lo desea puede resolverlos en PostgreSQL o MySQL.

1. Tome nota de la intercalación (collate) de la base de datos creada en el TP1.
2. Genere tres bases de datos adicionales, para lograr obtener una combinación de todas las variantes de sensibilidad a acentos y mayúsculas/minúsculas.
3. Cree un esquema denominado “ddbba” (por bases de datos aplicada). Todos los objetos que cree a partir de aquí deberán pertenecer a este esquema o a otro según corresponda. No debe usar el esquema default (dbo).
4. Recomendamos emplear la tabla y SP “registro” mencionados en el TP1 para operaciones de debugging.
5. Descargue de la web https://datos.gob.ar/dataset/otros-nombres-personas-fisicas/archivo/otros_2.1 todos los CSV de distintos períodos.
6. Genere la estructura de tabla para almacenar los datos descargados. Haga esto en las cuatro DB.
7. Importe los CSV descargados repartiéndolos en las cuatro DB. A los fines de esta práctica no es vital que lo reparta en cantidades precisamente iguales.
8. Es muy probable que algunos caracteres de acentos se hayan importado incorrectamente. Verifique y corrija lo necesario. Es importante que mantenga las letras acentuadas.
 - a. Hay muchas formas de lograr lo mismo. Puede por ejemplo buscar patrones de letras con el operador LIKE, o bien buscar valores específicos de caracteres. Preste atención al comportamiento en cada DB a la misma cláusula WHERE (por ejemplo) que utilice en el reemplazo.
9. Determine el primer y el último percentil de nombres más empleados (el 1% más empleado y el 1% menos empleado) para cada base de datos. Genere una vista en cada DB para dicha consulta.
10. Obtenga una lista de coincidencias entre el primer percentil entre las cuatro DB. Genere una vista para dicha consulta.
11. ¿Qué ocurre cuando compara cadenas de texto de distinta intercalación? ¿Cómo se resuelve esa situación?
12. Genere un SP que crea una tabla temporal global (usando doble numeral como prefijo de su nombre, puede emplear el esquema default) e inserta en la misma los

dos percentiles obtenidos antes. Repita lo mismo en las cuatro DB, utilice la misma tabla temporal global (puede generar el mismo SP en las cuatro DB).

13. Valiéndose de una CTE obtenga un resumen de los 100 nombres más utilizados en cada DB. Genere una vista con la consulta en cada DB.
14. Compare el contenido de la tabla temporal global con los 100 nombres más utilizados. Genere una vista con la consulta en cada DB.
15. ¿Qué ocurre cuando compara cadenas de texto en cada caso? ¿Qué intercalación le asignó el sistema a la tabla temporal global? ¿Qué puede ocurrir si implementa un sistema con una DB para la que no controla la intercalación global y decide usar tablas temporales globales como herramienta?
16. Verifique el resultado de las siguientes consultas:

```
select 1 Acierto where 'Cadena' = 'Cadena'
```

```
select 1 Acierto where 'CADENA' = 'Cádena' collate Modern_Spanish_CI_AI
```

```
select 1 Acierto where 'CADENA' = 'Cadena' collate Modern_Spanish_CS_AI
```

```
select 1 Acierto where 'CADENA' = 'Cádena' collate Modern_Spanish_CI_AS
```

¿Qué uso puede hacer de la cláusula COLLATE al comparar cadenas de texto (por ejemplo en búsquedas)?

Guarde periódicamente el script. Al finalizar genere un backup de la base de datos.