

# Proyecto final de la asignatura de bases de datos estructuradas

Ing. Fernando Arreola

Octubre 2024

## Objetivo

El alumno analizará un conjunto de información, generará su modelo de datos correspondiente, diseñará un proceso ETL para el procesamiento y resolverá una serie de consultas analíticas que permitan explotar la información, aplicando los conceptos vistos en el curso.

## Descripción del problema

El problema se divide en cinco partes:

### Análisis de información:

Los datos a emplear para el proyecto puede consultarlos en: <https://smn.conagua.gob.mx/es/web-service-api>. Revisar la documentación del sitio web y visualice una muestra de información. Comience a hacerse preguntas y buscar respuestas.

### Modelado de datos:

A partir de su análisis y de la información a procesar:

- Normalizar los datos
- Generar el *modelo relacional*
- Generar el modelo *entidad relación*
- Generar el modelo *físico*

### Procesamiento de datos:

Elaborar un flujo ETL en la herramienta *pentaho* que permita descargar los datos del sitio web por medio de una petición *GET* y realizar las siguientes transformaciones:

- Convertir de formato json a un formato estructurado
- Sustituir el valor numérico del atributo *ndia* por el día de la semana correspondiente
- Remover el atributo *raf*
- Asignar tipos de datos adecuados para que correspondan con los propuestos en sus modelos

Finalmente, cargar los datos en las tablas generadas como parte de su modelado de datos según corresponda.

### Analítica de datos:

Responder las siguientes preguntas/consultas:

- Mencione y explique detalladamente al menos dos ventajas y dos desventajas de la forma en que inicialmente se estaba representando la información.
- Generar una consulta que muestre los 5 municipios donde se vayan a presentar descensos de temperatura (*tmin*) mas marcados entre hoy y el siguiente día.
- Generar una consulta que permita obtener por estado, el municipio con la cuarta temperatura más alta.
- Si con los datos con los que está trabajando tuviera que resolver un problema o implementar alguna mejora de una situación real, ¿cuál podría ser un caso de estudio?

### Visualización

Finalmente, genere un *dashboard* que permita visualizar las *tmax* para un día en específico de cada municipio (al menos 15) por estado, ordenados de mayor a menor. El dashboard debe consumir los datos de su base de datos y tener una opción que permita seleccionar el estado y fecha de interés. Si se actualiza algún valor en la base de datos, debe reflejarse también en el dashboard.

## Entregables

Es necesario entregar lo que se solicita a continuación:

### Documento

Se deberá entregar un documento FORMAL elaborado en Latex, que contenga las siguientes secciones:

- **Introducción:** Breve descripción del análisis del problema, su(s) objetivos, su propuesta(s) de solución.
- **Plan de trabajo:** Queda a criterio de cada equipo el grado de detalle que se le dará a esta sección, pero mínimo debe contener descripción general de las actividades a realizar y su correspondiente plan de actividades, así como un breve detalle de qué hizo cada miembro del equipo.
- **Análisis:** Redactar un reporte que indique los hallazgos, ideas, limitaciones, etc. que se identificaron con base a los datos con los que se va a trabajar.
- **Diseño:** Descripción de lo realizado en las correspondientes fases de diseño de la base de datos, agregando los resultados de cada una de ellas.
- **Procesamiento:** Descripción detallada del funcionamiento de su proceso ETL.
- **Análítica de datos:** Explicación detallada de cada una de sus respuestas a cada pregunta/consulta solicitada.
- **Visualización:** Descripción de lo que hace su propuesta para visualizar el comportamiento de la información
- **Conclusiones:** Personales, detallando las dificultades, retos, aciertos, etc. que se presentaron en el proyecto.

## Código

A través de slack, se deberá entregar una carpeta por equipo donde se anexe lo siguiente, asignando nombres ADECUADOS para cada documento/archivo:

- El documento del punto anterior
- Script de creación de la base de datos y tablas
- Pipeline de datos
- Sentencias SQL para las consultas analíticas y el código generado para la etapa de visualización
- La presentación a emplear a la hora de exponer

## Exposición

Como parte de la evaluación del proyecto, se deberá realizar una presentación de su propuesta de solución. La idea es simular la presentación de un reporte de información a un directivo de una empresa, lo que implica explicación de los hallazgos, muestras de lo implementado para el análisis de información, solución de dudas, etc. Cada equipo es libre de decidir cómo hacer su presentación, el unico detalle es que debe ser TOTALMENTE formal.

## Puntos a considerar

Se tomará en cuenta lo siguiente, partiendo del porcentaje de calificación estipulado al inicio del semestre:

- Equipos que se copien en alguna parte del proyecto o que sus respuestas coincidan con sitios/herramientas web, su calificación del curso será 5.
- Equipo(s) que reprueben el desarrollo proyecto o la parte de exposición, deberán presentar examen final sin importar lo obtenido en los otros rubros del curso.
- La calificación del primer examen final será 30 % lo obtenido en el proyecto, 70 % lo obtenido en el examen.
- La seriedad, formalidad, presentación y calidad del proyecto, si bien no tendrán un valor numérico específico, pueden jugar a favor o en contra de la calificación de los rubros correspondientes.
- Cualquier punto no contemplado en el presente documento será determinado por el profesor.

## Recomendaciones

No es obligatorio, pero se sugieren las siguientes puntos:

- Delegar correctamente responsabilidades
- Empleo de herramientas colaborativas
- Empleo de arquitecturas cliente servidor remotas o servicios en la nube
- Redacción y presentación en inglés