

Universidad Pontificia Bolivariana  
Facultad Ingeniería en Energía, Computación y TIC  
Programas de Ingeniería en Sistemas e Informática – Ingeniería en Ciencia de Datos  
Curso de Tópicos Avanzados de Base de Datos  
NRC: 20635, 21333 – Periodo 202410

**Examen No. 2 – Valor 20%**  
**Índices y Mejoras en el Desempeño de Consultas**  
**Precipitaciones de lluvia en territorio antioqueño para el 2023**

**Fechas de Entrega**

- Calificación completa: viernes 22 de marzo de 2024, 6:00 pm (GMT -5)
- Calificación sobre 4.00: martes 26 de marzo de 2024, 6:00 pm (GMT -5)
- Cualquier entrega después de la hora especificada para calificación completa, entra en el rango de la calificación sobre **4.00 - sin excepciones.**
- Cualquier entrega después de la segunda fecha, su calificación será de **0.00 - sin excepciones.**
- Solo se responden dudas sobre el examen hasta el jueves 21 de marzo, 6:00 pm (GMT-5)

**Sobre supletorios**

**No se otorgarán plazos adicionales.** Recuerde que por indicaciones de la Escuela de Ingeniería, cualquier evaluación superior al 10% de la definitiva del curso que no sea presentado oportunamente, deberá tramitar autorización para ejecución de supletorio.

**Tiempo estimado de ejecución:** 8 horas

**Tipo de Evaluación:**

- Grupal, máximo 2 integrantes por equipo. Opcionalmente, puede presentarse individualmente.
- Los miembros del equipo de trabajo pueden ser de diferentes cursos (NRC) pero SIEMPRE de la misma materia.

**Nota importante sobre el trabajo grupal**

- Si lo hizo en parejas, **cada estudiante realiza el examen en un motor de base de datos diferente**, pero entregan el mismo diseño y las mismas consultas.
- Si lo hizo en parejas pero lo entrega individual, **se considera fraude.**
- Si lo hizo en parejas, no se hacen calificaciones de esfuerzos individuales. **Ambos sacan la misma nota.**

## Sobre el manejo del fraude

Por favor responda a la confianza que se le brinda con este examen, trabajando honestamente. Si tiene dificultades para cumplir con el proyecto, por favor realice una valoración **oportuna** de la situación para que trabajemos un plan de mejoramiento específico a sus necesidades. **¡No haga trampa!**

## Motores de base datos que se pueden utilizar en este examen:

- MySQL o MariaDB.
- PostgreSQL 16 o superior
- SQL Server 2022 o superior
- Oracle 21cXE – Oracle 23cFree

## Infraestructura:

El abastecimiento de infraestructura puede hacerse de la siguiente manera:

- Instalación local a través de máquina virtual, con sistema operativo de su predilección.
- Instalación local a través de contenedores, con distribución de Linux de su predilección.
- Instalación en nube, con el proveedor de nube de su preferencia, **siempre y cuando no genere costos.**

**Nota:** Se invita a **no realizar Instalación local directa sobre la estación de trabajo** del estudiante, como aplicación del sistema operativo base. Queremos fomentar el uso de tecnologías de contenedores o en nube como parte de las competencias profesionales.

## Origen de los datos

Los datos que se utilizarán en este examen fueron obtenidos de la plataforma de Datos Abiertos del Gobierno de Colombia, gestionados por el Departamento Administrativo Nacional de Estadísticas – DANE y el Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM, los cuales pueden accederse desde estas direcciones:

[https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about\\_data](https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about_data)

[https://www.datos.gov.co/Mapas-Nacionales/DIVIPOLA-C-digos-municipios/gdxc-w37w/about\\_data](https://www.datos.gov.co/Mapas-Nacionales/DIVIPOLA-C-digos-municipios/gdxc-w37w/about_data)

Se ha realizado las siguientes operaciones de filtrado y calidad de datos:

- Solo mediciones del año 2023.
- Solo mediciones en cabeceras de municipios en el departamento de Antioquia.
- Solo mediciones en milímetros (mm).
- Solo mediciones en los sensores de tipo 0240 – Precipitación
- Se normalizaron los nombres de municipios al valor utilizado en la Divipola del DANE.
- Se normalizaron los nombres de las estaciones para facilitar la legibilidad del registro.
- En la Divipola, solo se registran municipios del departamento de Antioquia.

Luego de este proceso, se tienen 575.551 mediciones para realizar las actividades de esta evaluación. Los archivos con los datos se denominan **ideam\_lluvias\_antioquia\_mediciones\_2023.csv** y **dane\_divipola\_antioquia.csv**

### **Formulación de Requerimientos:**

Dado un dominio de problema representado en los archivos de datos, se requiere construir un modelo de datos que cumpla con **la tercera forma normal en modelos relacionales (3FN)** y que pueda ser utilizado para **la valoración de mejoras de desempeño** que se obtendrán sobre las consultas propuestas.

**Los estudiantes propondrán 5 consultas** a las cuales se les realizará las acciones de mejoramiento. Cada una de las consultas deben **obligatoriamente incluir un componente de tiempo y ubicación**.

La actividad de valoración de desempeño busca generar el plan de ejecución e identificar reducciones en los costos de procesamiento o en la simplificación del árbol de operaciones utilizado por el motor de bases de datos para generar la consulta.

Para ello, se construirán dos esquemas: **inicial**, donde estarán los datos sin ningún proceso de normalización o de estructuración y el **corregido**, donde estarán los datos luego de realizarse el proceso de optimización.

Para asignar los códigos de los municipios, **se debe utilizar la información incluida en la definición de la Divipola**. Tener presente el tipo de datos que se utiliza allí para no perder información vital por definición de columna incorrecta.

Se busca que las consultas tengan **al menos una reducción del 10% en el costo de procesamiento. Si no se logra este umbral de mejoramiento, los estudiantes deben proponer una nueva consulta**.

Para lograr este cometido, los estudiantes realizarán estandarización de tablas y columnas, estandarización de tipo de datos, creación de claves primarias (PK) y claves foráneas (FK), creación de restricciones de unicidad (UK), creación de índices y eventualmente creación de vistas materializadas.

### **Nota importante:**

- Dos estudiantes individualmente o equipos de trabajo independientes **no pueden tener exactamente las mismas 5 consultas con exactamente las mismas columnas**. De presentarse esta situación, se entenderá que hubo **fraude**.
- Se entiende que es posible que el modelamiento resultado del proceso de normalización sea similar luego de la retroalimentación entregada en el examen No. 1, por lo que **no se considerará como fraude**.

### Entregables:

Los siguientes documentos deben entregarse como un archivo ZIP compartido a través de OneDrive y notificando vía correo electrónico al docente [juand.rodasm@upb.edu.co](mailto:juand.rodasm@upb.edu.co). **Cada entregable debe estar en archivo independiente. Por favor no compartir el documento en Microsoft Teams.**

- Diagrama relacional donde se evidencie las decisiones de diseño a nivel de nomenclatura de campos y tablas, definición de claves primarias y foráneas. El diagrama debe estar en imagen tipo JPEG o PNG.
- Script de implementación del modelo de datos, detallando la creación de la base de datos, el usuario, los esquemas, **asignación de privilegios mínimos**, creación de tablas, vistas, vistas materializadas y secuencias. El archivo tiene extensión .SQL
- Script de implementación de las consultas de SQL propuestas, **antes y después** del proceso de mejoramiento del desempeño. El archivo tiene extensión .SQL
- Documento en formato Microsoft Word, evidenciando el plan de ejecución antes y después para cada consulta. Incluir en él:
  - Capturas de pantalla antes y después de los cambios, se debe observar CLARAMENTE la reducción en el costo de procesamiento o en el número de operaciones requeridas.
  - Registros resultantes de la consulta antes y después de los cambios

**Si lo requiere, cambie el tamaño del papel utilizado para definir la página a unas dimensiones superiores. Igual, no será un documento impreso.** El archivo tiene extensión .DOCX

**En total son cuatro (4) documentos** que se deben incluir en el archivo ZIP para considerar la entrega como completa.

### Importante:

- En el correo electrónico de entrega del examen deben estar incluidos **TODOS** los miembros del equipo.
- Todos los entregables deben contener información de quienes lo está haciendo (Nombre completo así como ID de SIGAA). **Para la entrega en parejas, siempre debe tener el nombre de ambos miembros del equipo.**
- El correo electrónico **es la evidencia de la entrega oportuna de la solución del examen**. El docente siempre le dará confirmación de recepción y retroalimentación del trabajo presentado.

### Evaluables:

### **Etapas 1: Diseño e implementación del modelo de datos – 20%**

Con los datos suministrados en los archivos previamente indicados, implemente un modelo de datos que permita responder las preguntas propuestas por los estudiantes a través de consultas de SQL. De esta etapa salen los diagramas relacionales y los scripts de implementación del modelo de datos.

### **Etapas 2: Formulación de la pregunta e implementación de la consulta – 30%**

Formular 5 preguntas en lenguaje natural que incluyan característica de tiempo y ubicación; éstas deberán ser solucionadas a través de consultas SQL con exactamente los mismos resultados antes y después de la ejecución de las mejoras de desempeño. De aquí sale el script de implementación de las consultas de SQL.

### **Etapas 3: Evaluación de mejoras de desempeño – 50%**

Para cada una de las consultas formuladas en la etapa previa, realizar el registro de las mejoras de desempeño:

- Una imagen captura del “*explain plan*” antes y después de implementar las mejoras de desempeño.
- En el escenario antes, presente la consulta **SIN** la implementación de claves primarias, foráneas o índices de apoyo.
- En el escenario después, presente la consulta **CON** la implementación de claves primarias, foráneas y de índices de apoyo.

De aquí sale el documento en formato Microsoft Word.

### **Importante:**

**Se debe evidenciar mejora de al menos 10% en el costo total de ejecución de la consulta.** Consultas con mejora inferior a este valor o peor aún, con deterioro en el desempeño, no serán consideradas como entregas válidas.