

Weather Track: Sistema de Seguimiento y Gestión de Datos Meteorológicos y Alertas Meteorológicas

Ingeniería de datos

Presenta:

Cristian Camilo Ramos Moreno

Deiby Rafael Ospina Triviño

Profesora:

Luz Stella García Monsalve

E[scuela de Ingeniería, Ciencia y Tecnología](https://urosario.edu.co/escuela-de-ingenieria-ciencia-y-tecnologia?utm_source=Pauta&utm_medium=Google_Search&utm_term=Digital&utm_content=113&utm_urterm=701PX000001WibfYAC)

Universidad Del Rosario

Octubre,2024

Bogotá D.C

**Tercera Entrega**

***1.Visualización De Escenarios:***

* La visualización de escenarios se encuentra en un archivo llamado “graficas.py”,

Interfaz de usuario gráfica, Aplicación, Teams

Descripción generada automáticamente

***2.Ventajas Y Desventajas:***

***Análisis Cristian***

1. ***Función para generar alertas:***

* ***Ventajas:***
* *Elimina la necesidad de monitorear manualmente los datos para detectar condiciones anormales.*
* *Permite generar alertas de manera inmediata cuando se detectan eventos críticos.*
* *La lógica de generación de alertas está encapsulada en una función separada, lo que mejora la legibilidad y mantenibilidad del código.*
* ***Desventajas:***
* *A medida que se agregan más tipos de sensores y alertas, la función puede volverse más compleja y difícil de mantener.*
* *No hay mecanismos para personalizar las alertas según diferentes usuarios o aplicaciones.*
* *El código está estrechamente ligado a la estructura de la base de datos y podría requerir ajustes si se cambian las tablas o los campos involucrados.*

1. ***Evolución de la dirección del viento en cada ciudad:***

* ***Ventajas:***
* *La función agrupa los datos por ciudad, mes y año, proporcionando un resumen conciso de la evolución de la dirección del viento.*
* *Al utilizar RETURN QUERY, se puede integrar fácilmente en otras consultas o vistas, lo que permite realizar análisis más complejos.*
* *Puede manejar grandes conjuntos de datos de manera eficiente, gracias a las capacidades de agrupación y agregación de PostgreSQL.*
* ***Desventajas:***
* *La función se enfoca exclusivamente en la dirección del viento. Si se necesitan otros cálculos (velocidad, ráfagas, etc.), se requerirán funciones adicionales.*
* *Potenciales Problemas de Rendimiento: Para grandes conjuntos de datos, el cálculo del promedio puede ser costoso en términos de tiempo de ejecución, especialmente si no se utilizan índices adecuados.*
* *Falta de Manejo de Valores Atípicos: La función no incluye ningún mecanismo para identificar y manejar valores atípicos en los datos de la dirección del viento, lo que podría afectar la precisión de los resultados.*

1. ***Días con viento más fuerte en cada mes:***

* ***Ventajas:***
* *Utiliza ROW\_NUMBER() para identificar de manera eficiente el registro con la velocidad máxima dentro de cada partición (ciudad y mes).*
* *La función puede adaptarse fácilmente para encontrar otros valores extremos (mínimos, promedios) o para diferentes variables.*
* *Puede ser llamada desde otras consultas o aplicaciones para realizar análisis más complejos.*
* ***Desventajas:***
* *Se enfoca exclusivamente en la velocidad máxima. Para analizar otros aspectos del viento (como ráfagas o direcciones), se requerirían funciones adicionales.*
* *Si hay múltiples registros con la misma velocidad máxima en un día y ciudad, esta función solo devolverá uno de ellos de manera arbitraria.*
* *No proporciona información sobre si la velocidad máxima registrada es realmente un evento extremo o si es un valor normal para esa ciudad y época del año.*

1. ***Máximos y mínimos históricos de temperatura por ciudad:***

* ***Ventajas:***
* *Se usa SQL estructurado con subconsultas y particiones, lo que permite identificar fácilmente las diferentes partes del proceso (máximos, mínimos, uniones, etc.).*
* *Usa particiones (PARTITION BY) para manejar los datos agrupados por ciudad.*
* *Considera tanto temperaturas máximas como mínimas y sus fechas asociadas, garantizando que los valores sean precisos y relevantes.*
* ***Desventajas:***
* *Se enfoca exclusivamente en la velocidad máxima. Para analizar otros aspectos del viento (como ráfagas o direcciones), se requerirían funciones adicionales.*
* *Si hay múltiples registros con la misma velocidad máxima en un día y ciudad, esta función solo devolverá uno de ellos de manera arbitraria.*
* *No proporciona información sobre si la velocidad máxima registrada es realmente un evento extremo o si es un valor normal para esa ciudad y época del año.*

1. ***Promedio mensual de precipitaciones por ciudad:***

* ***Ventajas:***
* *Calcula el promedio mensual de precipitaciones, que es una métrica importante para análisis climáticos y toma de decisiones.*
* *La función es independiente y puede ser reutilizada en diversas aplicaciones, como análisis históricos, visualizaciones de datos o integraciones con sistemas externos.*
* *Es flexible y puede adaptarse fácilmente para otros parámetros climáticos siguiendo una estructura similar.*
* ***Desventajas:***
* *Si faltan datos de precipitaciones para ciertos meses o ciudades, los resultados pueden ser incompletos. Esto podría requerir manejar casos de datos faltantes explícitamente.*
* *A medida que aumenta la cantidad de datos históricos, la consulta podría volverse más lenta debido al cálculo en tiempo real de AVG y la agrupación.*

1. ***Total de alertas generadas por sensor:***

* ***Ventajas:***
* *La consulta es directa y fácil de entender, mostrando el total de alertas agrupadas por tipo de sensor.*
* *Permite identificar qué tipo de sensores generan más alertas, lo que puede ayudar a tomar decisiones sobre mantenimiento, priorización o calibración de dispositivos.*
* ***Desventajas:***
* *No incluye un desglose temporal, lo que limita la posibilidad de analizar tendencias o picos en alertas a lo largo del tiempo.*
* *Si las tablas alertas, sensores y tipo\_sensores contienen millones de registros, la agrupación puede volverse costosa en términos de tiempo de ejecución.*
* *No está diseñada para analizar grandes volúmenes de datos en tiempo real o en sistemas distribuidos.*

1. ***Promedio climático anual:***

* ***Ventajas:***
* *Calcula promedios anuales de tres parámetros clave (precipitación, temperatura y velocidad del viento), lo que facilita una visión climática completa por ciudad.*
* *Agrupar por año ayuda a identificar tendencias climáticas anuales y realizar comparaciones históricas.*
* *Combina datos de varios sensores por ciudad, promediando las mediciones para obtener una representación global del clima.*
* ***Desventajas:***
* *La función puede ser costosa si las tablas contienen un gran volumen de datos, ya que implica cálculos de agregados (AVG) y agrupaciones.*
* *La precisión de los resultados depende de que los datos estén completos y correctamente registrados. Faltas en alguna de las variables podrían afectar la calidad de los promedios.*

*>---------------------------------------------------------------------------------------------------------<*

***Análisis Deiby:***

***Ventajas y desventajas de las funciones SQL del proyecto***

1. ***Funciones de alertas*** 
   * ***Ventajas:***
     + Automatización del proceso de generación de alertas, reduciendo la intervención manual.
     + Asegura que las alertas sean consistentes y generadas en tiempo real al ingresar nuevos datos.
     + Facilita la detección temprana de situaciones climáticas críticas.
   * ***Desventajas:***
     + Puede ser costoso en términos de rendimiento cuando se manejan grandes volúmenes de datos.
     + La lógica de alertas es rígida y requiere modificar el código para ajustar umbrales o condiciones.
2. ***Evolución de la dirección del viento*** 
   * ***Ventajas:***
     + Permite analizar tendencias a lo largo del tiempo, útil para estudios meteorológicos.
     + Ofrece una vista compacta y resumida de los datos de viento.
   * ***Desventajas:***
     + No tiene en cuenta valores atípicos o patrones anómalos que podrían distorsionar los promedios.
     + Puede ser difícil interpretar solo con promedios, si no se acompañan de otras métricas.
3. ***Días con viento más fuerte*** 
   * ***Ventajas:***
     + Identifica eventos extremos de manera clara, útiles para análisis de riesgos.
     + Resalta patrones estacionales o mensuales en las velocidades máximas del viento.

* + ***Desventajas:***
    - No proporciona información contextual, como otras variables meteorológicas relacionadas.
    - Requiere ajustes si se cambia la definición de "viento más fuerte".

1. ***Máximos y mínimos históricos de temperatura*** 
   * ***Ventajas:***
     + Proporciona puntos de referencia útiles para análisis históricos y climáticos.
     + Fácil de interpretar y útil para reportes.
   * ***Desventajas:***
     + No incluye variaciones diarias ni contexto sobre los eventos máximos/mínimos.
     + Puede ser costoso para bases de datos muy grandes si no se optimizan índices.
2. ***Promedio mensual de precipitaciones*** 
   * ***Ventajas:***
     + Ideal para analizar patrones estacionales de lluvia.
     + Resalta variaciones a largo plazo en diferentes ciudades.
   * ***Desventajas:***
     + Puede ocultar eventos extremos de lluvia dentro del promedio.
     + Los datos faltantes o incompletos pueden sesgar los resultados.
3. ***Promedio anual climático*** 
   * ***Ventajas:***
     + Resume múltiples variables en una única consulta, facilitando comparaciones anuales.
     + Útil para identificar tendencias generales a largo plazo.
   * ***Desventajas:***
     + Percepción limitada de eventos atípicos o desviaciones a corto plazo.
     + Puede consumir recursos significativos en bases de datos grandes.

***Ventajas y desventajas de las visualizaciones***

1. ***Alertas por ciudad (gráfico de barras):***
   * ***Ventajas:***
     + Clara y fácil de interpretar.
     + Permite comparar directamente la cantidad de alertas entre ciudades.
   * ***Desventajas:***
     + No muestra detalles sobre el tipo de alerta o el tiempo en que ocurrieron.
     + Difícil de visualizar si hay muchas ciudades con valores similares.
2. ***Distribución de alertas por ciudad (gráfico circular):***
   * ***Ventajas:***

* Muestra proporciones relativas de forma clara.
* Ideal para identificar rápidamente qué ciudades generan más alertas.
  + ***Desventajas:***
    - Poco efectiva con muchas categorías, ya que las proporciones pequeñas son difíciles de diferenciar.
    - No muestra el contexto temporal ni los tipos de alertas.

1. ***Evolución de la dirección del viento (gráfico lineal):***
   * ***Ventajas:***
     + Excelente para identificar tendencias y patrones en el tiempo.
     + Claramente muestra cambios a lo largo del tiempo para cada ciudad.
   * ***Desventajas:***
     + Puede volverse confuso si hay muchas líneas en el gráfico.
     + No muestra la variabilidad diaria, solo promedios mensuales.
2. ***Días con viento más fuerte (gráfico de dispersión):***
   * ***Ventajas:***
     + Visualiza eventos extremos claramente.
     + Ideal para identificar patrones de eventos de viento fuerte en el tiempo.
   * ***Desventajas:***
     + Puede ser difícil de interpretar si los puntos están muy agrupados.
     + No incluye información de otras variables relevantes como dirección del viento.
3. ***Promedio mensual de precipitaciones (gráfico lineal):***
   * ***Ventajas:***
     + Útil para visualizar patrones estacionales y variaciones a largo plazo.
     + Fácil de comparar entre diferentes ciudades o períodos.
   * ***Desventajas:***
     + ***No muestra eventos de precipitación extremos ni el rango de valores.***
     + ***La línea puede ser engañosa si hay meses sin datos.***
4. ***Cantidad de sensores por tipo (gráfico de barras):***
   * ***Ventajas:***
     + Sencillo y fácil de interpretar para comparar cantidades absolutas.
     + Ideal para mostrar la distribución de sensores entre categorías.
   * ***Desventajas:***
     + No muestra información sobre cómo se utilizan los sensores.
     + Puede ser poco informativo si los valores son muy similares.
5. ***Alertas por tipo de sensor (gráfico circular):***
   * ***Ventajas:***
     + Proporciona una visión clara de qué tipos de sensores generan más alertas.
     + Fácil de entender para proporciones relativas.
   * ***Desventajas:***
     + No permite ver el detalle del volumen absoluto de alertas.
     + No contextualiza los datos temporalmente o por ubicación.
6. ***Comparación de velocidad máxima del viento (gráfico de barras):***
   * ***Ventajas:***
     + Ideal para destacar diferencias máximas entre ciudades o sensores.
     + Fácil de interpretar al comparar valores absolutos.
   * ***Desventajas:***
     + No incluye información sobre cuándo ocurrieron las velocidades máximas.
     + Puede volverse redundante si las diferencias son mínimas.

***3.Conclusiones:***

El desarrollo de este proyecto ha permitido a nuestro grupo adquirir experiencia práctica en el diseño e implementación de soluciones tecnológicas basadas en datos. A continuación, compartimos nuestras principales apreciaciones respecto a las diferentes etapas del proceso:

**Selección de fuentes de datos**

La selección de las fuentes de datos fue una etapa crucial que determinó la calidad y relevancia del proyecto. Optamos por el uso de datos meteorológicos en tiempo real y datos históricos de estaciones climáticas, lo cual nos permitió cubrir diversos escenarios de análisis. Sin embargo, enfrentamos desafíos relacionados con la limpieza y homogeneización de datos debido a inconsistencias y datos faltantes en las fuentes seleccionadas. Esto nos enseñó la importancia de evaluar exhaustivamente las fuentes antes de integrarlas al sistema.

* **Lecciones aprendidas:**
  + Es esencial contar con una fuente de datos confiable y bien documentada.
  + La preparación y limpieza de datos requiere tiempo y planificación.

**Carga de datos**

La carga masiva de datos fue un proceso complejo, especialmente al manejar grandes volúmenes de información. Implementamos estrategias de transformación previa y scripts de carga para automatizar el proceso. Aprendimos que la validación de datos antes de su ingreso a la base es esencial para evitar problemas futuros.

* **Lecciones aprendidas:**
  + Las pruebas en entornos controlados son críticas antes de cargar grandes volúmenes de datos.
  + Automatizar el proceso de carga reduce errores y agiliza la integración de nuevas fuentes.

**Creación de funciones y consultas SQL**

La creación de funciones y consultas SQL personalizadas fue un componente clave para el éxito del proyecto, permitiendo realizar análisis avanzados y optimizados sobre los datos meteorológicos. Las funciones facilitaron cálculos recurrentes, como promedios y máximos, mientras que las consultas específicas proporcionaron la flexibilidad necesaria para obtener información precisa y relevante. Este enfoque redujo la carga computacional y mejoró la eficiencia en el procesamiento de grandes volúmenes de datos.

**Lecciones aprendidas**

* El análisis de rendimiento de las consultas SQL fue esencial para garantizar tiempos de respuesta aceptables, especialmente al trabajar con conjuntos de datos masivos.
* Definir claramente las relaciones entre tablas y anticipar las necesidades de análisis permitió crear consultas más directas y funcionales.

**Desarrollo usando Dash**

Dash resultó ser una herramienta poderosa y flexible para la creación de aplicaciones interactivas. A través de esta plataforma, logramos integrar visualizaciones dinámicas que facilitaban el análisis de los datos meteorológicos. A pesar de su curva de aprendizaje inicial, Dash permitió un desarrollo rápido y una personalización significativa de las visualizaciones.

* **Lecciones aprendidas:**
  + Dash combina simplicidad y funcionalidad, siendo ideal para proyectos de análisis y visualización.
  + Es importante definir desde el inicio las necesidades de visualización para optimizar el desarrollo.

**Conexiones con la base de datos**

El desarrollo de conexiones entre Dash y la base de datos presentó desafíos relacionados con la eficiencia y la seguridad. Implementamos consultas optimizadas y estrategias de paginación para evitar la sobrecarga del sistema. Además, aprendimos a manejar conexiones concurrentes para garantizar un rendimiento estable.

* **Lecciones aprendidas:**
  + La optimización de consultas es crucial para garantizar tiempos de respuesta aceptables.
  + El manejo adecuado de credenciales y configuraciones de acceso es vital para proteger la base de datos.

**Apreciaciones finales**

El proyecto nos permitió integrar diversas competencias técnicas, desde la manipulación de datos hasta la creación de aplicaciones interactivas. Aunque enfrentamos múltiples desafíos, cada etapa nos brindó aprendizajes valiosos que fortalecieron nuestra capacidad para abordar problemas reales mediante soluciones tecnológicas. Este proceso reafirmó la importancia del trabajo en equipo y la planificación estratégica para el éxito en proyectos complejos.

En resumen, el proyecto no solo nos permitió aplicar los conocimientos adquiridos, sino que también destacó la relevancia de habilidades prácticas como la resolución de problemas, la adaptabilidad y la comunicación efectiva entre los miembros del equipo.