🚧 RETO 1. Análisis algorítmico de eventos de seguridad

Tipo de reto: Formativo
Unidad: Algoritmos de búsqueda y ordenamiento aplicados a ciberseguridad

1. Descripción del reto

En este reto, el estudiante desarrollará un módulo de análisis básico para eventos de ciberseguridad, utilizando algoritmos de búsqueda y ordenamiento sobre un conjunto de datos simulados.

El propósito es comprender cómo la elección de un algoritmo impacta directamente en el rendimiento de un sistema que procesa grandes volúmenes de información, como ocurre en los centros de operaciones de seguridad (SOC).

El estudiante deberá implementar, comparar y analizar diferentes estrategias algorítmicas, desarrollando criterio técnico para la toma de decisiones en escenarios reales.

2. Problema a resolver

Un sistema de monitoreo de seguridad recibe grandes volúmenes de eventos generados por accesos, conexiones de red y actividad de usuarios. Dentro de estos registros pueden existir eventos sospechosos que deben ser identificados de manera eficiente.

El estudiante deberá diseñar un módulo que permita buscar eventos sospechosos dentro de un conjunto de logs simulados, considerando volúmenes que pueden escalar hasta **1 000 000 de registros**.

El análisis debe incluir la comparación entre búsqueda lineal y búsqueda binaria, así como la necesidad de ordenar los datos previamente para optimizar las consultas.

3. Objetivos de aprendizaje

Al finalizar el reto, el estudiante será capaz de:

- Implementar algoritmos de búsqueda y ordenamiento en contextos de ciberseguridad
- Analizar la complejidad temporal de los algoritmos utilizados
- Comparar el rendimiento en distintos volúmenes de datos
- Justificar decisiones técnicas basadas en eficiencia y escalabilidad
- Comprender el impacto de estas decisiones en sistemas reales de monitoreo

4. Producto esperado

El estudiante deberá desarrollar un **cuaderno en Google Colab** que incluya:

- Generación o carga de datos simulados
- Implementación de algoritmos requeridos
- Ejecución de pruebas con distintos tamaños de entrada
- Medición de tiempos de ejecución
- Análisis técnico y conclusiones

5. Datos simulados requeridos

El conjunto de datos debe representar eventos de ciberseguridad e incluir al menos los siguientes campos:

- timestamp
- ip_origen
- usuario
- tipo_evento
- severidad
- estado

Tipos de eventos sugeridos

- login_exitoso
- login_fallido
- acceso_admin
- escaneo_puertos
- conexion_remota
- acceso_fuera_horario

Niveles de severidad

- baja
- media
- alta
- crítica

Condiciones

El estudiante deberá generar múltiples volúmenes de datos:

- 1 000
- 10 000

- 100 000
- hasta 1 000 000 registros

Los datos deben permitir consultas como:

- búsqueda de IP sospechosa
- eventos por timestamp
- eventos de alta severidad
- actividad repetida por usuario

6. Actividades obligatorias

6.1 Generación de datos

Construir o cargar un dataset coherente y estructurado.

6.2 Búsqueda lineal

Implementar búsqueda sin orden previo.

6.3 Algoritmo de ordenamiento

Implementar al menos uno:

- QuickSort
- MergeSort

6.4 Búsqueda binaria

Aplicar búsqueda binaria sobre datos ordenados.

6.5 Comparación experimental

Medir tiempos de ejecución en distintos tamaños de entrada.

6.6 Análisis técnico

Responder:

- ¿Cuándo vale la pena ordenar los datos antes de buscar?
- ¿Qué ocurre si los datos llegan en tiempo real?
- ¿Qué algoritmo es más eficiente y por qué?
- ¿Cómo escala la solución?

7. Uso de herramientas de IA generativa

Se **recomienda el uso de herramientas de IA generativa** como apoyo para la **generación de datos simulados**, siempre que el estudiante mantenga control sobre la calidad de la información.

El uso de IA puede aplicarse para:

- generar logs realistas
- simular escenarios
- escalar volúmenes de datos

⚠️ Importante:
La implementación de algoritmos, análisis y conclusiones deben ser **propias del estudiante**.

8. Declaración obligatoria de uso de IA

El uso de herramientas de IA implica una **obligación de transparencia académica**.

Requisito obligatorio:

El estudiante debe incluir una sección de **ANEXOS** con:

- Todos los **prompts utilizados**
- Respuestas relevantes generadas por IA
- Descripción del propósito de cada prompt

Ejemplo:

Prompt:
"Genera 10 000 logs de ciberseguridad con IPs y timestamps"

Uso:
Creación de dataset para pruebas

⚠️ La omisión de esta información será considerada falta académica.

9. Requisitos técnicos

- Desarrollo obligatorio en **Google Colab**
- Código claro, estructurado y comentado
- Evidencia de pruebas con distintos volúmenes
- Análisis de complejidad (Big-O)
- Comparación entre algoritmos

10. Formato de entrega

El estudiante deberá entregar un **PDF exportado desde Google Colab** que incluya:

- Portada

- Planteamiento del problema
- Descripción del dataset
- Implementación
- Resultados
- Análisis
- Conclusiones
- Anexos (uso de IA)

11. Rúbrica de evaluación (100 puntos)

Criterio	**Puntaje**
Comprensión del problema	10
Calidad de datos simulados	15
Búsqueda lineal	10
Algoritmo de ordenamiento	15
Búsqueda binaria	15
Comparación experimental	15
Análisis y justificación técnica	15
Presentación y organización	5
Uso responsable de IA	5

12. Evidencia de logro esperada

El estudiante no solo debe implementar algoritmos, sino demostrar que:

- comprende su comportamiento
- puede comparar alternativas
- justifica decisiones técnicas
- entiende el impacto en sistemas reales de ciberseguridad

Pregunta de cierre

Si tu sistema recibe eventos en tiempo real, ¿seguirías usando búsqueda binaria?

Justifica en términos de **estructura de datos y complejidad**.