

Quiz 5-6

1. Un índice no clusterizado es aquel en el que las filas de una tabla no están organizadas físicamente en el mismo orden que las claves del índice. En su lugar, el índice no clusterizado mantiene una lista de las direcciones de las filas que contienen la clave de búsqueda. Un índice no clusterizado con INCLUDE permite agregar columnas adicionales al índice, esto evita la necesidad de buscar la tabla de datos subyacente. La principal diferencia entre ambos es que un índice no clusterizado no organiza físicamente los datos, mientras que un índice clusterizado sí lo hace.
2. El concepto de "memory footprint" se refiere al tamaño de memoria requerido para almacenar un conjunto de datos en un sistema informático. Al crear índices, un alto memory footprint puede ralentizar el rendimiento del sistema, ya que esto puede afectar la capacidad del sistema para realizar otras tareas. Además, un alto memory footprint puede llevar a la paginación a disco, lo que significa que parte de la información se guarda en el disco duro en lugar de en la memoria principal del sistema.
3. Lo que recomendaría sería crear índices en las columnas country, city y date para mejorar la velocidad de búsqueda, esto evitaría el ineficiente scan sobre todos los datos. Debido al alto volumen de datos almacenados una buena opción sería usar una base de datos NoSQL. Además estas bases son buenas para manejar datos no estructurados como son los documentos JSON de payload que incluyen los eventos. Usar MongoDB sería recomendable debido a que este funciona con documentos BSON (Binary JSON) para almacenar los datos. BSON es muy similar a JSON. MongoDB además permite crear índices en campos específicos dentro de los documentos BSON, lo que facilita la búsqueda de información en los documentos de la columna payload.