

# SGY y DATA-SET

## Aplicando mmap



# Br. Brayan Ceballos

## Archivo SGY

Un archivo SGY es un formato de archivo que se usa en la industria de la exploración sísmica para guardar datos sísmicos. Estos archivos tienen información que se ha obtenido por equipos sísmicos, que sirven para estudiar las capas subterráneas de la Tierra y encontrar petróleo, gas u otros recursos naturales. También tienen datos sísmicos que se han obtenido usando técnicas de adquisición sísmica, como la reflexión sísmica, que consiste en enviar ondas de sonido al subsuelo y medir los tiempos que tardan las ondas reflejadas en volver a la superficie. Estos datos se guardan en un formato de archivo binario especial que facilita su procesamiento y análisis posterior. Se usan comúnmente en la industria del petróleo y el gas para la exploración y producción de hidrocarburos. Los datos sísmicos que hay en los archivos SGY se usan para crear imágenes de las capas subterráneas de la Tierra y para detectar posibles yacimientos de petróleo y gas.

## Data-Set SGY

El data set está compuesto por las siguientes secciones: Textual Header, Binary Header y la traza.

- **Textual Header:** Al leer los primeros 3200 bits del archivo se encuentra información adicional sobre los datos sísmicos que hay en el archivo. A diferencia del Binary Header (que se hablará más adelante), que tiene información en un formato binario especializado, el Textual Header tiene información en un formato de texto legible. Puede tener información sobre el proyecto de adquisición sísmica, el equipo que se usó para obtener los datos, los parámetros de procesamiento y análisis de los datos, la ubicación geográfica de la adquisición sísmica y otros metadatos relevantes. Esta información puede ser útil para comprender el contexto en el que se obtuvieron los datos sísmicos y para asegurarse de que se procesen y analicen correctamente.
- **Binary Header:** Después de leer los 3200 bits anteriores del Textual Header, ahora se leen 400 bits del Binary Header, que contiene información sobre la configuración del equipo de adquisición sísmica que se usó para obtener los datos, como el número de

canales, la frecuencia de muestreo, la duración de la adquisición y el tamaño de cada traza. También puede tener información sobre la ubicación geográfica de la adquisición sísmica, la profundidad de la capa terrestre en la que se hizo la medición y otros metadatos relevantes. La información que hay en el Binary Header es crítica para la correcta interpretación de los datos sísmicos que hay en el archivo SGY. Los programas de procesamiento y análisis de datos sísmicos usan esta información para interpretar los datos y crear imágenes de las capas geológicas subterráneas.

- **Traza:** Después de leer el Textual Header y el Binary Header, que fueron 3600 bits en total, se leen a continuación 240 bits, donde se guarda la información de la traza. Las trazas son los datos sísmicos que se han obtenido en un punto de adquisición específico. Cada traza representa una medición de la reflexión de una onda sísmica en un punto determinado debajo de la superficie terrestre.

### Librerías utilizadas

- **stdio.h:** Biblioteca estándar para entrada y salida de datos.
- **stdlib.h:** Biblioteca estándar para funciones generales de memoria y utilidad.
- **sys/mman.h:** Biblioteca para mapeo de archivos en memoria.
- **sys/stat.h:** Biblioteca para información de archivos.
- **fcntl.h:** Biblioteca para control de archivos.
- **unistd.h:** Biblioteca para funciones generales de Unix.

### Explicando el código

- **Inclusión de librerías:** Las primeras líneas del código incluyen las bibliotecas necesarias para las distintas funcionalidades que se implementan.
- **Definición de constantes:** Se define la constante `TRACE_HEADER_SIZE` para especificar el tamaño del encabezado de la traza en un archivo SEG Y (240 bytes).
- **Función principal:** La función `main()` contiene la lógica del programa.
- **Apertura del archivo:** Se abre el archivo "arch.sgy" en modo lectura utilizando la función `open()`. Si la apertura falla, se imprime un mensaje de error y se termina la ejecución.

- **Obtención del tamaño del archivo:** Se utiliza la función `fstat()` para obtener el tamaño total del archivo abierto.
- **Mapeo del archivo en memoria:** Se utiliza la función `mmap()` para mapear el archivo completo en la memoria del sistema. Esto permite acceder a los datos del archivo de forma más rápida y eficiente que con la lectura tradicional.
- **Salto del encabezado de la traza:** Se calcula la posición del inicio de la traza sumando el tamaño del encabezado al puntero del archivo mapeado.
- **Lectura de la traza:** Se utiliza un bucle `for` para recorrer cada muestra de la traza y acceder a su valor utilizando el puntero actualizado.
- **Visualización de la traza:** Se imprime el número de la muestra y su valor correspondiente con dos decimales de precisión.
- **Desmapeo del archivo:** Se libera la memoria asociada al archivo mapeado utilizando la función `munmap()`.
- **Cierre del archivo:** Se cierra el archivo abierto con la función `close()`.
- **Retorno exitoso:** Se retorna el valor 0 para indicar la ejecución exitosa del programa.

### Ventajas del uso de Mmap

- **Mejora el rendimiento:** El acceso a los datos del archivo en memoria es más rápido que la lectura tradicional del disco.
- **Reduce el desgaste del disco:** El acceso a los datos se realiza sin necesidad de operaciones de lectura/escritura en el disco.
- **Posibilita el acceso aleatorio:** Se puede acceder a cualquier muestra de la traza fácilmente con un cálculo de la posición en memoria.

### Desventajas del uso de Mmap

- **Mayor consumo de memoria:** El archivo completo se carga en la memoria del sistema, lo que puede ser un factor limitante en sistemas con poca RAM. Posibilidad de corrupción de datos: Si el proceso se interrumpe antes de desmapear el archivo, los datos podrían quedar corrompidos.

## **Conclusión**

Este código demuestra una forma eficiente de leer y visualizar datos sísmicos almacenados en archivos SEGY utilizando C y Mmap. Al comprender las funcionalidades de las librerías utilizadas y el funcionamiento del código, se puede adaptar y ampliar para realizar diferentes tareas de procesamiento y análisis de datos sísmicos.