

Objetivos

Construir una librería para manipular y analizar datos en dos dimensiones. Tiene que contemplar la posibilidad de extender las funcionalidades.

Alcance

El lenguaje para desarrollar esta propuesta es Java. La carga y descarga de datos se puede realizar a través de archivos CSV o estructuras en memoria.

Descripción de alto nivel del sistema

Vamos a desarrollar una librería en Java orientada a la manipulación y análisis de datos tabulares, que permitirá realizar operaciones como selección, filtrado, agregación y ordenamiento de datos. El diseño se centrará en ser extensible y mantenible, garantizando una buena modularidad para incorporar nuevas funcionalidades de manera sencilla a largo plazo.

Requisitos Funcionales

- La librería debe proporcionar información resumida sobre cada instancia.
- Las columnas deben tener un tipo de dato asociado y debe tener un valor especial en caso de faltante.
- La librería debe proveer acceso indexado a nivel de fila y columna.
- La librería debe soportar lectura y escritura entre instancias en memoria y el formato CSV en disco.
- La librería debe presentar en formato texto información sobre la instancia en forma de tabla.
- La librería debe permitir generar tablas:
 - Cargando de un CSV
 - Con copia profunda desde otra instancia de la misma librería
 - Con una estructura de dos dimensiones nativa de Java
 - Con una secuencia lineal nativa de Java
- La librería debe permitir modificar la tabla:
 - Accediendo directo a una celda y asignando un nuevo valor
 - Insertando una columna nueva a partir de otra columna (con misma cantidad de elementos que filas)
 - Insertando una columna nueva a partir de una secuencia lineal nativa de Java (con misma cantidad de elementos que filas)
 - Eliminando una columna
 - Eliminando una fila
- La librería debe permitir seleccionar de forma parcial parte de la tabla, eligiendo ciertas columnas y/o ciertas filas.
- La librería debe permitir seleccionar parte de la tabla pero aplicando un filtro según los valores de las celdas.
- La librería debe permitir una copia profunda de una tabla.
- La librería debe generar una concatenación de dos tablas que posean misma cantidad de columnas, con mismos tipos de datos.

- La librería debe permitir ordenar la estructura tabular según un criterio
- La librería debe permitir rellenar los valores faltantes.
- La librería debe mostrar valores aleatorios de filas según un porcentaje del total de la estructura.

Requisitos No Funcionales

- La librería debe procesar operaciones en menos de 5 segundos para conjuntos de datos de hasta 100,000 filas.
- La librería debe usar un máximo de 50 MB de memoria por cada 100,000 filas con 10 columnas.
- La librería debe ser capaz de manejar un crecimiento de hasta 10,000 filas adicionales sin que el tiempo de procesamiento aumente en más del 20%.
- La librería debe ser capaz de manejar un crecimiento de hasta 5 columnas adicionales sin que el tiempo de procesamiento de un conjunto de datos de 100,000 filas aumente en más del 25%.
- La librería debe garantizar una disponibilidad continua durante la ejecución del programa, con un tiempo de inactividad máximo permitido de 1 segundo por cada 24 horas de operación
- La librería debe ser compatible con versiones del entorno Java desde la 8 en adelante.
- La librería debe funcionar en los siguientes sistemas operativos:
 - Windows:
 - Windows 11(64-bits)
 - Windows 10(64-bits)
 - Windows 7 (64-bits)
 - MacOS:
 - macOS Ventura (13)
 - macOS Monterey (12)
 - macOS Big Sur (11)
 - Linux:
 - Ubuntu 20.04 LTS y superiores
 - Debian 11 Bullseye y superiores
 - Fedora 36 y superiores
- La librería debe proporcionar mensajes de error claros y útiles que faciliten la identificación y resolución de problemas.
- La librería debe incluir documentación detallada y ejemplos de uso claros para facilitar una comprensión adecuada y asegurar su correcto uso.
- El código debe ser modular, bien estructurado y documentado para facilitar su mantenimiento y evolución.
- La librería debe garantizar una probabilidad menor al 0.001% de que los datos no se corrompan durante las operaciones de lectura, escritura o modificación.
- La librería debe ser capaz de manejar errores de manera robusta, recuperándose de situaciones imprevistas críticas en menos de 2 minutos y en caso de errores menores en menos de 15 segundos.