



Práctica 01: Procesamiento por Lotes 1

Maestro:

Javier Rosales Martinez

Materia:

Seminario de Solución de Problemas de Sistemas Operativos

Sección:

D06

Alumno:

Alejandro Covarrubias Sánchez

Código:

221350192

Antecedentes

El procesamiento por lotes es una técnica que consiste en agrupar y ejecutar una serie de tareas o datos en bloques, llamados lotes, en lugar de procesarlos uno a uno de forma individual e inmediata.

Este programa busca leer un archivo de texto inicial que contiene una gran cantidad de datos línea por línea en lotes, para dar formato a cada una y luego escribir el contenido resultante en un nuevo archivo de salida.

Procesar los datos en lotes de esta manera permite manejar grandes volúmenes de información de manera más eficiente, reduciendo la carga en la memoria y optimizando el rendimiento. Al dividir el procesamiento en bloques, se evita tener que cargar en memoria y manipular todo el archivo de una sola vez, lo que es particularmente beneficioso cuando se trabaja con archivos grandes. Este enfoque también facilita la escritura más estructurada en el archivo de salida y puede mejorar la velocidad del programa al reducir el número de operaciones de entrada y salida.

Metodología

El programa fue escrito en Python y utiliza únicamente librerías predeterminadas, como las funciones `open()` para la manipulación de archivos y estructuras de control básicas como `try` y `except` para manejar excepciones. No se utilizaron librerías externas.

Se tienen dos funciones principales: `process_line()`, que procesa cada línea individualmente convirtiendo números hexadecimales a decimales y viceversa y proporciona el formato de salida a cada una, y `read_and_process_file()`, que lee el archivo en lotes de un tamaño determinado, procesa cada línea y luego guarda el resultado en un archivo de salida.

```

def process_line(line):

    # Procesar la línea y realizar conversiones

    prefix, suffix = line.strip().split('/', 1)

    hex_parts = prefix.split(':')

    decimal_numbers = [int(number, 16) for number in hex_parts]

    parts = suffix.strip().split(',')

    second_string = parts[2]

    decimal_numbers_from_suffix = parts[5].split('.')

    hex_numbers = [format(int(number), 'X') for number in
                    decimal_numbers_from_suffix]

    # Crear el formato de salida

    decimal_numbers_str = ' : '.join(map(str, decimal_numbers))

    hex_numbers_str = '.'.join(hex_numbers)

    output_line = f"{second_string} : {decimal_numbers_str} :
                    {hex_numbers_str}\n"

    return output_line

```

El algoritmo clave es el procesamiento por lotes, que permite manejar archivos grandes de manera eficiente al procesar grupos de líneas a la vez. Además, utiliza conversiones de formatos numéricos (hexadecimales y decimales) para transformar los datos adecuadamente.

Conclusión

El programa actual a entregar está diseñado para leer y procesar archivos en lotes de 100 líneas, formateando datos específicos de cada línea para escribirlos en un archivo de salida. Actualmente, la funcionalidad básica está completa y operativa, incluyendo la conversión de datos y el manejo de errores.

Se podría mejorar optimizando el uso de memoria y el tamaño de los lotes, especialmente para archivos extremadamente grandes. Además el programa asume que el archivo de entrada está previamente formateado de manera consistente, por lo que también se puede mejorar agregando más validaciones de datos para manejar posibles inconsistencias en la entrada. También se podría considerar la incorporación de herramientas para el procesamiento paralelo o asíncrono para mejorar el rendimiento en escenarios con archivos muy grandes.

Referencias

Amazon Web Services. (n.d.). *¿Qué es el procesamiento por lotes? - Explicación de los sistemas de procesamiento por lotes.* AWS.

<https://aws.amazon.com/es/what-is/batch-processing/>