Cursos Obligatorios



Procesamiento de datos masivos

- 1. Nombre y apellidos del estudiante: Alex Anthony Prieto Romani
- 2. Tiempo empleado en entender el defecto del programa: 1 día (un fin de semana)
- 3. Descripción del defecto:
 - a) Circunstancias bajo las que falla el programa:

El programa falla cuando se ejecutan múltiples Combiners. Específicamente, cuando los datos de un cliente se dividen entre dos o más Combiners, puede resultar en una pérdida de información sobre las tiendas visitadas.

b) Motivos por los que falla el programa:

El defecto principal está en el Combiner. Este está diseñado para emitir solo hasta 3 tiendas por cliente, independientemente de cuántas tiendas únicas haya visitado realmente el cliente. Cuando los datos se dividen entre múltiples Combiners, cada uno emite un máximo de 3 tiendas, lo que puede resultar en que el Reducer reciba menos de 3 tiendas únicas para un cliente que realmente visitó 3 o más tiendas.

- c) Directrices para corregir el defecto:
- El mapper el error que encontré es que no maneja adecuadamente las líneas de entrada vacías o mal formateadas, lo que podría causar un error de ValueError si la línea no tiene dos partes separadas por una tabulación.
- Modificar el Combiner para que emita todas las tiendas únicas que recibe para cada cliente; cuando cambia el cliente y ha acumulado al menos 3 tiendas únicas, imprime solo las primeras 3 tiendas únicas visitadas por ese cliente. Esto esto implica en eliminar la limitación de 3 tiendas en el Combiner.
- Asegurarse que el Reducer pueda manejar correctamente la entrada del Combiner modificado, manteniendo la lógica de contar las tiendas únicas y emitir el cliente si visitó 3 o más tiendas, el error ahí es que la indentación es inconsistente en varios lugares, lo que causaría errores de sintaxis en Python.
- Para todos los casos, se les agrego una función como buena práctica, con el fin de mantener el código modular y reutilizable; además de realizar la condicional de ejecución para asegurar que la función "mapper()", "combiner()" o "reducer()" se ejecuta solo cuando el script se ejecuta directamente y no cuando se importa como un módulo.
- Se cambió #!/usr/bin/python3 por #!/usr/bin/env python3, ya que es más flexible por que busca el intérprete de Python en el entorno del usuario y no directamente la ruta al intérprete de Python.

Universidad Internacional de Valencia

Cursos Obligatorios

Procesamiento de datos masivos

- 4. ¿Te fue útil la información de depuración (la imagen reduccion.jpg y que el defecto se encontraba en >1 Combiners) para entender el defecto? Sí, la información de depuración fue muy útil por las siguientes razones:
 - La imagen reduccion.jpg proporcionó un ejemplo claro y conciso de cómo el programa puede fallar cuando se utilizan múltiples Combiners.
 - La información sobre que el defecto se encuentra en >1 Combiners dirigió inmediatamente la atención al comportamiento del Combiner y cómo interactúa con múltiples instancias.
 - Estos datos permitieron reproducir el problema con un conjunto mínimo de datos, facilitando el análisis y la comprensión del defecto.
 - Sin esta información, habría sido más difícil identificar la causa raíz del problema, ya que el programa funciona correctamente en configuraciones más simples con un solo Combiner.