**Ejemplos de transformaciones.**

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

En este ejemplo se utiliza **filter** para encontrar todos los números que son divisibles por 2 dentro del RDD, luego se utiliza **map** para elevar al cuadrado cada uno de los elementos resultantes de **filter**. Finalmente se usa **collect** para mostrar los elementos.

Imagen que contiene Texto

Descripción generada automáticamente

Una mejora del ejemplo seria aplicar una condición dentro del **map** para que automáticamente eleve al cuadrado todos los números pares, de lo contrario, devuelve None. Y finalmente el **filter** se encarga de eliminar todos los None.

Texto

Descripción generada automáticamente

En este ejemplo se utilizar **flatMap** para dividir las oraciones en palabras, señalando que cada palabra esta separada por un espacio. Luego aplicamos **distinct** para no mostrar palabras repetidas.

Texto

Descripción generada automáticamente

Para optimizar este ejemplo se utiliza **mapPartitions**, la cual es muy útil cuando se trabaja con datos en bloques mas grandes, puesto que ejecuta la función en toda la partición, evitando así las llamadas repetitivas con las que trabaja **map**.

Texto

Descripción generada automáticamente

En este ejemplo se utilizar **union** para unir los datos de dos RDD en una sola lista, de igual manera con **intersection**, para ver los elementos en común de ambos RDD.

Texto

Descripción generada automáticamente

Para mejorar el ejemplo anterior, deberíamos eliminar los elementos repetidos dentro de cada RDD usando **distinct**, así nuestro proceso sería mucho más eficiente al momento de trabajar con una gran cantidad de datos.

**Ejemplos de acciones**

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

En este ejemplo usamos **reduce** para calcular la suma de todos los elementos del RDD, también usamos **collect** para mostrar todos los elementos en una lista.

Texto

Descripción generada automáticamente

Una mejora de este ejemplo seria aplicar un filtro a los elementos del RDD y luego trabajar con los datos que cumplan con el filtro.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

En este ejemplo utilizamos **count** para contar los elementos de un RDD y **first** para obtener el primer elemento.

Imagen que contiene Texto

Descripción generada automáticamente

Una alternativa seria trabajar con un filtro que me separe a las palabras que tengan más de 4 letras, luego contar cuantas de estas hay en el RDD, y después, seleccionar a la primera palabra que cumpla con la condición y escribirla en mayúsculas.

Texto

Descripción generada automáticamente

En este ejemplo, usamos **take** para obtener los 3 primeros elementos del RDD, también usamos **max** y **min** para obtener el mayor y el menor valor respectivamente.

Texto

Descripción generada automáticamente

Una mejora seria aplicar un **sortBy** antes de aplicar **take** para obtener los valores de forma ordenada.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

En este ejemplo, **countByKey** contará las veces que aparece cada clave dentro del RDD.

Texto

Descripción generada automáticamente

Una versión mejorada seria usar **countByKey** para contar las veces que un producto se vendió y luego calcular el promedio de ventas por cada producto.