

Universidad Central de Venezuela

Facultad de Ciencias

Escuela de Computación

Introducción a la Ciencia de los Datos

Multipliación matricial haciendo uso de la función Map y Reduce

Para empezar se desarrollaron los scripts, `productmv.R`, `productmm.R`, `MapReduce.R`, `indexes.R`, `memlimit.R`, y un script para realizar pruebas `test.R`. La descripción de dichos archivos se encuentra a continuación.

`Indexes.R`: La función `indexes` recibe como parámetros la dimensión de la matriz y el número de iteración en el cual nos encontramos, y esta va a retornar los índices a los cuales debemos acceder secuencialmente para poder extraer la fila correspondiente a la iteración, del archivo `.csv`.

`MapReduce.R`: El script `MapReduce.R` contiene dos funciones, La función `map` que dado 2 numeros esta retornará el producto entre ellos. Y la función `reduce` que dado una matriz, retornará un vector columna con la suma de las filas de la matriz en las posiciones correspondientes, esto se hace de tal forma de ir armando la matriz mediante vectores columna.

`Memlimit.R`: El script `memlimit.R` contiene dos funciones que se usan en caso que estemos en un producto Matriz-Vector o Matriz-Matriz, estas funciones reciben como parámetros la dimensión de la matriz y la cantidad de bytes, y estas retornarán la cantidad de filas que deben ser leídas en el `.csv`, téngase en cuenta que se considera que la entrada no es un string sino números enteros separados por comas, entonces se considera que el tamaño del `int` es de 1 byte.

`Productmv.R`: La función `productmv` es una interfaz la cual recibe la ruta de los archivos `.csv` que contienen la matriz y el vector a multiplicar, la dimensión resultante y la cantidad de bytes a usar, esta función llamará a otra llamada `productmv_vector`, que recibe los `dataFrame` que se almacenan al leer los `csv`, la dimensión y los bytes a utilizar.

La función `productmv_vector` tratará de recorrer el archivo como si fuese una matriz, sin embargo como el acceso no es secuencial en el archivo, se utiliza una función explicada anteriormente que retorna las posiciones necesarias haciendo uso de una formula matemática. Luego se procede a llamar a la función `map` con cada una de estas posiciones del `data frame` y su correspondiente del vector, los resultados serán almacenados en un vector fila y a su vez cada uno de estos vectores

filas, almacenados en un vector columna para luego llamar a la función reduce con esta matriz resultante. Al finalizar este proceso se obtiene el producto Matriz-Vector.

$$\begin{array}{cccc} 1 & 2 & 3 & 6 \\ 4 & 5 & 6 & \rightarrow 15 \\ 7 & 8 & 9 & 24 \end{array}$$

Ejemplificación de la función Reduce

Productmm.R: El producto Matriz-Matriz es simplemente realizar N veces el producto Matriz-Vector explicado anteriormente, por lo que la función productmm simplemente aplicará N veces la función productmv_vector y los vectores columnas retornados por la función reduce, los almacenará en un vector fila para armar la matriz deseada.

Integrantes:

- Akchar, Jean
- Rodríguez, Cesar
- Romero, Daniel

Caracas, 9 de Abril de 2016