|  |
| --- |
| Ejercicios de MapReduce |

|  |  |  |
| --- | --- | --- |
| **1.** | Sea la Matriz M: [(1,2,3,4);(5,6,7,8);(9,10,11,12);(13,14,15,16)] (x filas) Y el vector V: [1,2,3,4] Se quiere calcular usando MapReduce el producto MxV. Indicar cuales de los siguientes pares son generados por los procesos Map. |  |

|  |  |  |
| --- | --- | --- |
|  | a) | (3,9) |
|  | b) | (2,8) |
|  | c) | (2,70) |
|  | d) | (4,16) |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | | Answer submitted:   **a)** | |  | | You have answered the question correctly. | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2.** | Un website tiene informacion sobre sus usuarios en un archivo distribuido de la forma:   |  |  |  | | --- | --- | --- | | **user\_id** | **name** | **registered\_on** | | 1 | Juan Perez | 2013-10-21 | | 2 | Jorge Gomez | 2013-7-11 |   Por otro lado registra los clicks que los usuarios hacen en las paginas del sitio en otro archivo distribuido de tipo:   |  |  | | --- | --- | | **user\_id** | **section** | | 1 | sports | | 2 | news | | 1 | news | | 1 | news | | 1 | news | | 1 | news | | 2 | news | | 2 | sports |   Para una campaña de marketing se desea conocer los clientes que hicieron click al menos 10 veces en la seccion sport y que se han registrado en el sitio antes del 14 de Enero del 2013.  Luego de realizar en MapReduce el proceso pedido se pide analizar cual de las siguientes afirmaciones es correcta: |  |

|  |  |  |
| --- | --- | --- |
|  | a) | Si se usan combiners el proceso reduce no es necesario |
|  | b) | La clave de reduce es simple (tiene un campo) |
|  | c) | Es necesario hacer 2 iteraciones de Map y Reduce |
|  | d) | Es necesario hacer dos procesos Map pero solo un Reduce |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | | Answer submitted:   **d)** | |  | | You have answered the question correctly. | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **3.** | El archivo "R" tiene 2 campos A y B y sus contenidos son:   |  |  | | --- | --- | | A | B | | 0 | 1 | | 1 | 2 | | 2 | 3 |   El archivo S tiene 2 campos B y C y sus contenidos son:   |  |  | | --- | --- | | B | C | | 0 | 1 | | 1 | 2 | | 2 | 3 |   Si queremos calcular el JOIN de ambos archivos usando MapReduce x el campo B indicar cuales de los siguientes seran entrada de los procesos Reduce. |  |

|  |  |  |
| --- | --- | --- |
|  | a) | (1, [(1,(R,0)), (1,(S,2))]) |
|  | b) | (1, [(S,2)]) |
|  | c) | (2, [(R,1), (S,3)]) |
|  | d) | (1,(S,2)) |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | | Answer submitted:   **c)** | |  | | You have answered the question correctly. | |

|  |  |  |
| --- | --- | --- |
| **4.** | Un proceso MapReduce recibe números enteros (las claves no importan). El proceso Map factoriza el número "n" en sus divisores primos (Ej: 39 = 3,13) y genera pares (p,n) (la clave es el divisor). Reduce toma cada factor primo y suma los numeros enteros que son divisibles por el mismo). reduce(p, [i1, i2, ...,ik]) is (p,i1+i2+...+ik) Calcular la salida si los numeros a procesar son: 15,21,24,30,49. Identificar cuales de los siguientes pares es parte de la salida final del proceso MapReduce. |  |

|  |  |  |
| --- | --- | --- |
|  | a) | (5,45) |
|  | b) | (7,48) |
|  | c) | (6,54) |
|  | d) | (7,119) |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | | Answer submitted:   **a)** | |  | | You have answered the question correctly. | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **5.** | Se tiene un archivo con dos campos: ALFA y BETA   |  |  | | --- | --- | | **ALFA** | **BETA** | | A | B | | B | C | | B | A | | A | D | | A | E | | C | B |   Se quiere realizar un proceso MapReduce que devuelva todos los valores de ALFA sin duplicados. ¿Cuales de las siguientes afirmaciones es correcta? |  |

|  |  |  |
| --- | --- | --- |
|  | a) | La unica funcion de Reduce es eliminar resultados repetidos |
|  | b) | Solo hace falta usar un proceso Map |
|  | c) | No es necesario usar Reducers |
|  | d) | Hacen falta 2 procesos Map |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | | Answer submitted:   **a)** | |  | | You have answered the question correctly. | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **6.** | Se tiene un archivo que almacena links entre paginas web de la forma url1=>url2   |  |  | | --- | --- | | CNN | Amazon | | Amazon | CNN | | Amazon | Google | | Amazon | Facebook | | Google | CNN | | Google | Twitter | | Facebook | Google |   Se quiere realizar en MapReduce un proceso que devuelva todos los caminos de 3 paginas posibles de la forma X,Y,Z tales que hay un link de X a Y y otro de Y a Z. ¿Cuales de las siguientes afirmaciones son correctas? |  |

|  |  |  |
| --- | --- | --- |
|  | a) | Hay en total 9 caminos de longitud 3 |
|  | b) | Es necesario hacer 2 pasadas de MapReduce |
|  | c) | Hay en total 10 caminos de longitud 3 |
|  | d) | Una de las salias de Map es (CNN,Google) |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | --- | | Answer submitted:   **c)** | |  | | You have answered the question correctly. | |

Estamos buscando algo de tipo url1,url2,url3 donde existe el link url1,url2 y url2,url3

Este tipo de ejercicios a veces salen bien si pensamos primero en que necesitariamos en el reducer.

En este caso lo que necesita reducer es al elemento del medio (url2) como clave y a algo que le diga

que cosas aparecen a izquierda (1) y derecha (2) de esa url.

Entonces si un reduce recibe algo de tipo (A,[(B,1),(C,2),(D,2),(E,1)])

puede generar las triplas: BAD BAC EAC y EAD

A partir de esto es mucho mas facil pensar lo que tiene que hacer map.