**Chapter 12 test problems**

1. Write SQL queries on the below database table that return:

1. All the temperature data.
2. All the cities, but without repetition.
3. All the records for India.
4. All the Fall records.
5. The city, country and season for which the average rainfall is between 200 and 400 millimeters.
6. The city and country for which the average Fall temperature is above 20 degrees, in increasing temperature order.
7. The total annual rainfall for Cairo.
8. The total rainfall for each season.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **City** | **Country** | **Season** | **Temperature (C)** | **Rainfall (mm)** |
| Mumbai | India | Winter | 24.8 | 5.9 |
| Mumbai | India | Spring | 28.4 | 16.2 |
| Mumbai | India | Summer | 27.9 | 1549.4 |
| Mumbai | India | Fall | 27.6 | 346.0 |
| London | United Kingdom | Winter | 4.2 | 207.7 |
| London | United Kingdom | Spring | 8.3 | 169.6 |
| London | United Kingdom | Summer | 15.7 | 157.0 |
| London | United Kingdom | Fall | 10.4 | 218.5 |
| Cairo | Egypt | Winter | 13.6 | 16.5 |
| Cairo | Egypt | Spring | 20.7 | 6.5 |
| Cairo | Egypt | Summer | 27.7 | 0.1 |
| Cairo | Egypt | Fall | 22.2 | 4.5 |

2. Write list comprehension statements that use the range function and generate the following lists:

1. The list of all integers in the range from 0 to 100.
2. The list of all multiples of 7 in the range from 0 to 100.
3. The list of squares of all multiples of 7 in the range from 0 to 100.
4. The list containing lists, with one list for every multiple of 7 in the range from 0 to 100; the list corresponding to each multiple i of 7 should be [i-1, i, i+1].

3. Suppose list words is defined as follows:

>>> words = ['The', 'quick', 'brown', 'fox', 'jumps', 'over', 'the', 'lazy', 'dog']

Write list comprehension expressions that use list words and generate the following lists:

1. ['THE', 'QUICK', 'BROWN', 'FOX', 'JUMPS', 'OVER', 'THE', 'LAZY', 'DOG']
2. ['the', 'quick', 'brown', 'fox', 'jumps', 'over', 'the', 'lazy', 'dog']
3. [3, 5, 5, 3, 5, 4, 3, 4, 3] (the list of lengths of words in list words).
4. [['THE', 'the', 3], ['QUICK', 'quick', 5], ['BROWN', 'brown', 5], ['FOX', 'fox', 3], ['JUMPS', 'jumps', 5], ['OVER', 'over', 4], ['THE', 'the', 3], ['LAZY', 'lazy', 4], ['DOG', 'dog', 3]] (the list containing a list for every word of list words, where each list contains the word in uppercase and lowercase and the length of the word.)
5. ['The', 'quick', 'brown', 'fox', 'jumps', 'over', 'the', 'lazy', 'dog'] (the list of words in list words containing 4 or more characters.)

4. Suppose that the MapReduce framework is used to compute the frequency of every word in the sentence

"Bison from Buffalo, that bison from Buffalo bully, themselves bully bison from Buffalo"

Write down the output of the map step, the intermediate partition step, and the reduce step.

5. Let lst be a list containing lists of numbers such as, for example,

[[2,4,5], [3,6,7,8], [1,2,3,7], [5,7,9]]

You would like to use the MapReduce framework to construct a list that contains, for every unique number appearing in lst, a tuple mapping the number to the list of indexes of lists in lst that contain the number. For the list above, the result would be

[(2,[0,2]), (3,[1,2]), (4,[0]), (5,[0,3]), (6,[1]), (7,[2,3]), (8,[1]), (9,[3])]

Write the mapper and reducer functions for this problem.

6. The running time taken by executing a sequential algorithm is 34.13 seconds and the running time taken by a parallel algorithm on the same input is 8.17 seconds. What is the speedup achieved by the parallel algorithm? What is the minimum number of CPUSs or CPU cores required to achieve this speedup?