

Post and comment building on Stackoverflow¹

Structured to Hierarchical pattern (Challenge 5_a)

For this exercise you have to prepare a report including:

- Your data collections comment describing what they represent.
- Your source codes: a file precising in a comment the compiling and execution commands.
- Propose a test battery and report execution measures.
- Compress everything and sent a .zip or a .tar to genoveva.vargas@gmail.com

1.1 Problem statement

Problem: Given a list of posts and comments, create a structured XML hierarchy to nest comments with their related post. Given the output, perform a self-join operation to create a question, answer, and comment hierarchy.

1.2 Implementation

Look at page 76 of the book “Map Reduce design patterns” and see the proposed `Map` and `Reduce` codes. Prepare a data collection of your choice and implement the solution. As in previous challenges, you can use StackOverflow or any other source for generating your collection. Program the first part of the example.

- Explain the why it was necessary to program the `Driver`. Describe the principle of this program. Which is the role of the `MultipleInputs` object? Use examples for this purpose.
- Comment the use of `nestElements`.
- Explain the principle of having two mapper classes? Does this strategy have specific impact to the performance of the solution?
- Prepare collections of different sizes to run your tests trying to get to the limits of your solution.
- Make comparisons. Do not hesitate to prepare graphics.

Look at page 80 of the book “Map Reduce design patterns” and see the proposed `Map` and `Reduce` codes for programming the second part of the example. In this example, we use the comments associated with the posts to associate the post answers with the post questions.

- Explain the difference of use of the same pattern for addressing both problems.
- Explain the principle of self-join and how it is implemented with your solution.
- How does the mapper distinguish between a question and an answer record?

¹ This challenge is an example proposed in the book MapReduce design patterns, pp. 76.