

Distinct user IDs¹

Filtering pattern (Challenge 4_d)

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 user's comments, determine the distinct set of user IDs.

1.2 Implementation

Look at page 68 of the book "Map Reduce design patterns" and see the proposed Map and Reduce codes.

- Prepare your input collection. Describe its content and its provenance.
- Explain the "strategy" of feeding records with "null" values to the reducer.
- Propose and implement a combiner.
- Test your implementation with different data collection sizes and discuss results.

¹ This challenge is an example proposed in the book MapReduce design patterns, pp. 68.