

RANDOM SHARDS

Input splits:

key: category name

value: review

Mapper:

increasing counter;

when counter = max. number of shards:

reset to 0

replace key by counter value

Input:

key: random category name (a number)

value: review

Reducer:

saves each category in its own file