

8 Algorithms (fms27)

An e-commerce website sells an average of m items per second, from a catalogue of n digital items. Among other details the website keeps track, for every sale, of a timestamp, the item's code and the sale price. It maintains three bestseller lists, refreshed every s seconds:

- the k best selling items ever, ranked by how many units were sold;
 - the k best selling items of the past 30 days, ranked by how many units were sold;
 - the k highest-revenue items of this calendar year, ranked by total revenue since 1st of January of this year.
- (a) Each item will be held in a record. Describe all the data structures that must refer to these records to implement the required functionality. Describe all the fields that the record must have to implement the required functionality, and how each of these fields has to be updated and when. [5 marks]
- (b) The obvious baseline solution is to re-sort the n items and to take the top k every time the bestseller lists must be produced. Assuming the number of items and the given average rates stay constant, what is its asymptotic worst-case time cost per unit time? [1 mark]
- (c) Describe three alternative strategies, each better than the baseline, to implement the required functionality. Use the heaps or search trees of your choice, explaining precisely what you would store in each data structure to implement the required functionality. Describe, in each case, how to initialize the data structures, how to update the data structures after each sale, how to recompute the three bestseller lists every s seconds, together with the worst-case asymptotic time cost of each operation as a function of m, n, k, s (cost per unit time for the second and third operations). [6 marks]
- (d) Recommend the most appropriate strategy for $m = 10^4, n = 10^9, k = 10^2, s = 10^0$, with justification. [2 marks]
- (e) Repeat for $n = 10^{14}$ and the other parameters unchanged. [3 marks]
- (f) Repeat for $m = 10^{-4}, n = 10^2, k = 10^1, s = 10^5$, reasoning about the structural difference between this and the websites of cases (d) and (e). [3 marks]