examly

Derive an algorithm to find difficulty of a question given the following details

1) Question type -- MCQ, Fillup, Programming, Match the following
2) Manually assigned difficulty -- Easy, Medium, Hard
3) Total number of students who have attended this question
    a) Time taken by each student to answer this question
    b) Number of times the answer was changed if it is MCQ type question
    c) Number of times the program was compiled if it is programming question
    d) Number of hints used
    e) Programming language used if it is programming question (c, c++ , java etc)
4) Feedback given for this question by other students
5) Total number of students who have answered it right
6) Total number of students who have answered it wrong
7) Total number of students who have answered it partially correct
8) Maximum marks allocated for this question

The main goal of this exercise is to come up with a scalable solution. The total number of questions will be in the range of 100000 to 1 million. Total number of students will be in the range of 1 million to 10 million. The total number of question attempts in the worst case will be 30 million.
Assume all possible corner cases with respect to the question and question attempt data (Eg - Question was not attempted by anyone, the same student has attempted the same question multiple times with increasing or decreasing accuracy)

The following is expected as part of the solution

1) Algorithm as pseudocode or formula
2) Explanation and proof for the Algorithm in plain easy to understand english
3) Create the input data on your own for the volume required in the performance metrics specified below
4) Code in github as a standalone API which will accept the input data as file and print out the difficulty (in programming language of your choice)
5) Performance run metrics for
    a) 100 question 100 attempts
    b) 10000 question 10000 attempts
    c) 100000 question for 1 million attempts