

Description

Solution

Discuss (999+)

Submissions

It can be easily stated that at first we have to count the frequencies of each character in the string. After change the equal frequencies so that all become unique. But doing this in the most optimised way and process which will be discussed here.

- So as stated first step is to find the frequencies of the characters the string. Which can be done using vector/ array of size 26. (As there are 26 characters in alphabet which can be in the string). Time Taken = $O(26)$ which is constant i.e. $O(1)$.
- Now we can't increase the frequency of any character in the string as only deletion is allowed. So sort frequencies from greatest to smallest. So sort the array. Time Taken = $O(26 \log 26)$ which is constant now = $O(N)$.

Note : We have to only deal with the frequencies not the characters. So position of the frequencies matter as we will never need which character's frequency are dealt with.

- Now traverse the array from right to left as rightmost element is the greatest. We don't need to check element. So start traversing from 24th position. (As 25th is the last element). Time taken = $O(25)$ so total time is $O(N)$.
- Now we will continue till frequencies become 0 for any element in the array as other elements left break at the position where frequency is 0.
- If the current frequency (`freq[i]`) is equal or greater than the previous one (`freq[i+1]`) then we make current frequency less than previous one but it should be also greater than or equal to 0. (Frequency can't be 0). So `freq[i] = max(0, freq[i+1] - 1)`.

In case of doubt : So here a question may arise that why so much headache as it is sorted `freq[i]` and `freq[i+1]` just do `ans++` and reduce the frequency by 1 i.e. `freq[i]--`. Yes in many cases it will work but if more than 2 frequencies are equal then if we decrease a frequency then in next iteration the frequency of the previous one and if we only reduce 1 then the answer will not be correct. So we make the frequency of the previous one and check the difference.

Example : If frequencies be [2,2,2,2] then if we just decrease the freq by 1 in each step then next step frequencies will be [2,2,1,2], del = 1 -> [2,1,1,2], del = 2 -> [1,1,1,2], del = 3 which is not the desired result. If difference is taken the result becomes [2,2,2,2] -> [2,2,1,2], del = 1 -> [2,0,1,2], del = 2 -> [1,0,1,2], del = 3 -> [0,0,1,2], del = 4 -> [0,0,0,2], del = 5 which is the desired ans.

- But we also need the previous value of the frequency so that we can calculate the number of frequency characters deleted. So we store the previous freq as `prev` before the step mentioned before.
- So we add the reduced freq to the ans as `del += prev - freq[i]` (del is the answer).
- Return `del` at the end which will contain the total number of frequency decreased i.e. total characters deleted.

Example :

Let the string be : abbccdddeeffffggg

So count of a=1, b=2, c=2, d=3, e=2, f=4, g=3. So freq = {1,2,2,3,2,4,3,0,...,0}

So after sorting the elements of the freq array are = {0,...,0, 1,2,2,2,3,3,4}