362. Design Hit Counter

Medium Design Queue Array Hash Table Binary Search Leetcode Link

Problem Description

The problem at hand is designing a system that keeps track of the number of hits (or actions/events) that occur on a resource in the last 5 minutes. Importantly, the system is expected to handle hits in a chronological order, meaning that the timestamps for these hits are always increasing. The design should consist of a hit counter that can accept a new hit at a specific timestamp, with timestamps measured in seconds, and retrieve the total count of hits that occurred in the past 5 minutes from any given timestamp.

To summarize, the `HitCounter` class should provide two primary functionalities:

1. `hit(int timestamp)`: Record a new hit at the given timestamp.
2. `getHits(int timestamp)`: Fetch the number of hits in the last 300 seconds (5 minutes) from the given timestamp.

This counter could be used, for example, on a website to track the number of hits to a server, on an application to monitor the number of occurrences of a particular event, or any similar use-case where such time-bound counting is required.

Intuition

When considering how to design the `HitCounter`, one must realize that while we need to keep track of hits, storing every single hit with its timestamp is not efficient, especially as time progresses. We only care about hits in the last 5 minutes, not beyond that. Therefore, we must handle hits in such a way that avoids unnecessary memory use and ensures efficient retrieval.

The intuition behind the solution is to use a data structure that can record the hits while allowing fast insertion and look-up. A Python `Counter` object can handle the accumulation of hits well since it uses a hash table under the hood. Each hit is associated with a timestamp, and this timestamp acts as key in the `Counter`. The corresponding value is the number of hits that occurred at that timestamp.

The `hit` method updates the counter for the given timestamp. Multiple hits at the same timestamp simply increment the count for that timestamp.

For the `getHits` method, the goal is to calculate the sum of all the hits that happened from `(timestamp - 300)` to `timestamp`, because these represent the hits within the last 5 minutes. The method iterates through all items in the `Counter`, but only sums those where the timestamp of the hit is within the last 300 seconds from the given `timestamp`.

In practice, the `Counter` can potentially grow large if hits aren't removed after they become irrelevant (older than 5 minutes). To improve the implementation, one might periodically clean the `Counter` by removing old hits from it, to keep the memory footprint small and the `getHits` method efficient. However, the provided solution does not implement this optimization, and it's something users of the class would need to handle separately if needed.

Solution Approach

The implementation of the `HitCounter` class uses two primary methods as part of its API: `hit` and `getHits`.

The `Counter` data structure from Python's standard library collection is employed here to keep a tally of the hits. A `Counter` is essentially a dictionary or a hash map where keys are the items to be counted and the values are the counts. This is especially useful for our use case because it allows constant-time operations for storing and updating the number of hits per timestamp.

Let's delve into the implementation details of both methods:

###Hit Method When the `hit` method is called with a timestamp, it performs the following operations:

1. It checks if the timestamp is already present as a key in the `Counter`.
2. If it is, it increments the count associated with that timestamp.
3. If not, it creates a new entry with the timestamp as the key and sets the count to 1. Here's the snippet of the code that handles the hit:

```
1 def hit(self, timestamp: int) -> None:
2     self.counter[timestamp] += 1
```

The `+=` operation here is the Pythonic way to increment the value for a given key in a `Counter`. It's efficient because it does not require a search for the key before incrementation as the `Counter` handles that internally.

###GetHits Method The `getHits` method retrieves the count of hits within the past 5 minutes from a given timestamp and it operates as follows:

1. It initializes a variable to hold the sum of hits.
2. It iterates over all items in the counter.
3. For each item, it checks if the timestamp of that item is within the past 300 seconds of the given current timestamp.
4. If it is, it adds the count of hits for that timestamp to the sum.
5. After iterating through all the items, it returns the sum. The following snippet from the code explains this logic:

```
1 def getHits(self, timestamp: int) -> int:
2     return sum(t for t, v in self.counter.items() if t + 300 > timestamp)
```

Here, we use a list comprehension to iterate over the items of the `Counter`, summing up the values for the keys that meet the time condition (i.e., within the last 300 seconds). This method is straightforward and functional, but as mentioned earlier, it can become inefficient if the `Counter` grows too large due to old entries not being removed.

Overall, the pattern used in the solution is quite effective for this specific problem. However, in a real-world scenario where time efficiency for the `getHits` method becomes crucial, further optimization could be considered, like cleaning up old timestamps at regular intervals or after a certain number of hits to maintain the `Counter` size.

Example Walkthrough

Let's walk through a small example to illustrate how the `HitCounter` operates. Let's assume the following sequence of hits and queries:

1. A hit at timestamp 10.
2. A hit at timestamp 20.
3. A hit at timestamp 30.
4. A query for the number of hits at timestamp 40.

Now, let's see how the `HitCounter` will handle these events:

Recording hits

- We call `hit(10)`. The `counter` is updated to `{10: 1}`.
- We call `hit(20)`. The `counter` is updated to `{10: 1, 20: 1}`.
- We call `hit(30)`. The `counter` is updated to `{10: 1, 20: 1, 30: 1}`.

At each hit, the counter records the timestamp with a count. If there were multiple hits at the same timestamp, the count would increase accordingly; however, in this example, there is only one hit per timestamp.

Query for hits

- Next, we call `getHits(40)`. We need to sum counts of hits happening from timestamp `40 - 300 = -260` (but since a timestamp cannot be negative, we effectively start at 0) to 40.
- The `counter` has three timestamps entries: 10, 20, and 30, all of which fall within the last 300 seconds (5 minutes) from timestamp 40.
- Thus, the `getHits(40)` sums up the count of hits at all current timestamps and returns `1 + 1 + 1 = 3`.

The count of hits within the last 5 minutes from timestamp 40 is three, as there are no hits outside the 5-minute window yet. With this approach, querying for hits at any point provides the total number of hits in the preceding 5 minutes in an efficient manner, taking into account only the relevant timestamps.

Python Solution

```
1 from collections import Counter
2
3 class HitCounter:
4     def __init__(self):
5         """
6         Initialize the HitCounter with a Counter to keep track of timestamp occurrences.
7         """
8         self.hits = Counter()
9
10     def hit(self, timestamp: int) -> None:
11         """
12         Record a hit at a given timestamp.
13         Each hit increments the count for the specific timestamp.
14
15         :param timestamp: The current timestamp (in seconds granularity).
16         """
17         self.hits[timestamp] += 1
18
19     def get_hits(self, timestamp: int) -> int:
20         """
21         Retrieve the number of hits in the past 5 minutes (300 seconds) from the current timestamp.
22
23         :param timestamp: The timestamp at which to get the number of hits.
24         :return: Total number of hits in the last 5 minutes.
25         """
26         # Filter out the timestamps that are older than 5 minutes and sum their hit counts.
27         return sum(count for time, count in self.hits.items() if time > timestamp - 300)
28
29 # Example usage
30 # hit_counter = HitCounter()
31 # hit_counter.hit(timestamp)
32 # total_hits = hit_counter.get_hits(timestamp)
33
```

Java Solution

```
1 import java.util.HashMap;
2 import java.util.Map;
3
4 public class HitCounter {
5     // Use a map to store the count of hits for each timestamp
6     private Map<Integer, Integer> hitCounts;
7
8     /** Constructor to initialize the HitCounter object. */
9     public HitCounter() {
10         hitCounts = new HashMap<>(); // Initialize the map for storing hits
11     }
12
13     /**
14      * Record a hit at the given timestamp.
15      *
16      * @param timestamp - The current timestamp in seconds granularity.
17      */
18     public void hit(int timestamp) {
19         // Increment the hit count for the given timestamp, if it exists; otherwise, set it to 1
20         hitCounts.put(timestamp, hitCounts.getOrDefault(timestamp, 0) + 1);
21     }
22
23     /**
24      * Return the number of hits in the past 5 minutes.
25      *
26      * @param timestamp - The current timestamp in seconds granularity.
27      * @return the total number of hits in the last 5 minutes.
28      */
29     public int getHits(int timestamp) {
30         int hits = 0; // Variable to accumulate the number of hits
31
32         // Iterate through all entries in the map
33         for (Map.Entry<Integer, Integer> entry : hitCounts.entrySet()) {
34             // Check if the entry's timestamp is within the past 5 minutes from the given current timestamp
35             if (timestamp - entry.getKey() < 300) { // 300 seconds equals 5 minutes
36                 // Sum up the hits that are within the past 5 minutes
37                 hits += entry.getValue();
38             }
39         }
40         // Return the total number of hits
41         return hits;
42     }
43 }
44
45 // Usage example
46 // HitCounter hitCounter = new HitCounter();
47 // hitCounter.hit(timestamp);
48 // int numberOfHits = hitCounter.getHits(timestamp);
49
```

C++ Solution

```
1 #include <unordered_map>
2
3 class HitCounter {
4     // Use an unordered map to store the count of hits for each timestamp
5     private:
6         std::unordered_map<int, int> hitCounts;
7
8     public:
9         /** Constructor to initialize the HitCounter object. */
10         HitCounter() {
11             // The map is automatically initialized when the object is created
12         }
13
14         /**
15          * Record a hit at the given timestamp.
16          *
17          * @param timestamp - The current timestamp in seconds granularity.
18          */
19         void hit(int timestamp) {
20             // Increment the hit count for the given timestamp
21             // If it does not exist, it is inserted into the map with a count of 0, then incremented
22             hitCounts[timestamp]++;
23         }
24
25         /**
26          * Return the number of hits in the past 5 minutes.
27          *
28          * @param timestamp - The current timestamp in seconds granularity.
29          * @return the total number of hits in the last 5 minutes.
30          */
31         int getHits(int timestamp) {
32             int hits = 0; // Variable to accumulate the number of hits
33
34             // Iterate through all entries in the map
35             for (auto &entry : hitCounts) {
36                 // Check if the entry's timestamp is within the past 5 minutes from the current timestamp
37                 if (timestamp - entry.first < 300) { // 300 seconds equals 5 minutes
38                     // Sum up the hits that are within the past 5 minutes
39                     hits += entry.second;
40                 }
41             }
42             // Return the total number of hits
43             return hits;
44         }
45 };
46
47 // Usage example
48 // HitCounter hitCounter;
49 // hitCounter.hit(timestamp);
50 // int numberOfHits = hitCounter.getHits(timestamp);
51
```

Typescript Solution

```
1 // A map to store the count of hits for each timestamp
2 let hitCounts: Map<number, number> = new Map<number, number>();
3
4 /**
5  * Record a hit at the given timestamp.
6  *
7  * @param timestamp - The current timestamp in seconds granularity.
8  */
9 function hit(timestamp: number): void {
10     // Increment the hit count for the given timestamp, or set it to 1 if not present
11     const currentCount = hitCounts.get(timestamp) || 0;
12     hitCounts.set(timestamp, currentCount + 1);
13 }
14
15 /**
16  * Return the number of hits in the past 5 minutes.
17  *
18  * @param timestamp - The current timestamp in seconds granularity.
19  * @return the total number of hits in the last 5 minutes.
20  */
21 function getHits(timestamp: number): number {
22     let hits = 0; // Variable to accumulate the number of hits
23
24     // Iterate through all key-value pairs in the map
25     hitCounts.forEach((count, time) => {
26         // Check if the time is within the past 5 minutes from the current timestamp
27         if (timestamp - time < 300) { // 300 seconds equals 5 minutes
28             hits += count; // Add up the hits within the past 5 minutes
29         }
30     });
31
32     // Return the total number of hits
33     return hits;
34 }
35
36 // Usage example
37 // hit(timestamp);
38 // let numberOfHits = getHits(timestamp);
39
```

Time and Space Complexity

Time Complexity

`init` ($$\mathcal{O}(1)$$): Initializing the counter data structure is a constant time operation.

`hit` ($$\mathcal{O}(1)$$): Incrementing the counter for a given timestamp is a constant time operation assuming that the underlying data structure has $$O(1)$$ access time. The `Counter` in Python often provides $$O(1)$$ time complexity for insertion and access.

`getHits` ($$\mathcal{O}(N)$$): This operation involves iterating over all entries in the counter and summing the values for timestamps within the last 5 minutes. In the worst case scenario, $$N$$ represents the number of unique timestamps that have been recorded. The complexity becomes $$O(N)$$ when we need to iterate through all timestamps. However, depending on the use case, if timestamps are in a dense range and we get hits at almost every second, the actual time complexity may approach $$O(300)$$ since we are interested only in the hits in the last 5 minutes (300 seconds). Nevertheless, the worst-case scenario remains $$O(N)$$.

Space Complexity

The space complexity is $$\mathcal{O}(N)$$, where $$N$$ is the number of unique timestamps recorded. This is because the counter needs to store each individual timestamp's hit count. In a scenario where the system runs indefinitely, the counter can grow unbounded as more unique timestamps are recorded. The memory usage becomes a function of the number of unique timestamps, thus causing a linear growth in space complexity.