**Bayoan Heredia**

**Problem:** Calculating genomic coverage from next-generation sequencing data

**Solution Approach:**

reads.cvs – contains 2millions data rows

loci.csv – contains 1000 data rows

That’s means that there 2billions of iterations which this could take a lot of time.

I implemented a collection (hashtable) to reduce the reads data. There 2million of data in that file, therefore there a high probability that some data in that file can be repeated.

The result of that [start, length] will contain a value with how many times that [start, length] was found.

Then the loci will check on that collection and instead of reading through the same [start, length] just add the number stored on the collection.

A possible implementation to make it faster is with parallel programing where the read files will be run depending the cores of the computer and at the end collect each result.

**Bonus:**

Using the brute force (two for loop) will be 2billions of iterations that takes about 10minutes.

My solution I reduced the number of reads and reduced the time to less than a 1minute**.**