Step 3: Compute User Session Information (Optional Bonus Question)

Question 3.b: The maximum query session length

- Answer: 82,091 sec
- Main algorithm: The UDF GetSessionLength converts *hour* field into timestamp in unit of second and sets a user's list of session lengths. That is, it is in charge of imposing cut off where appropriate. The fixed session length of 15 minutes (900 seconds) was used. The main loop to set each user's session lengths is as follows:

```
int lptr = 0;
int rptr;
for(rptr=lptr+1;rptr<timestamp.size();rptr++) {
    int difference = timestamp.get(rptr)-timestamp.get(rptr-1);
    if(difference>=cutofflen) {
        sessionlength.add(timestamp.get(rptr-1)-timestamp.get(lptr)+cutofflen);
        ptr = rptr;
        rptr = lptr+1;
    }
    else if(difference<cutofflen && rptr==timestamp.size()-1) {
        sessionlength.add(timestamp.get(rptr)-timestamp.get(lptr)+cutofflen);
        lptr = timestamp.size()-1;
    }
    else if(difference<cutofflen) {
        continue;
    }

    if(lptr==timestamp.size()-1) {
      sessionlength.add(cutofflen);
      break;
    }
}
```

There are two timestamp pointers: left pointer (*lptr*) and right pointer (*rptr*). *lptr* always points the start point of a session. As *rptr* increases, when the difference between timestamps is larger than or same to 15 minutes, cut off is imposed. After that, the timestamp, which *rptr* is pointing, becomes the starts of another session.

Question 3.c: The average query session length

- Answer: 1,253.2808 sec

Question 3.d: The standard deviation of query session lengths

- Answer: 2499.599867978873
- Main algorithm: The UDF GetSquaredDiff returns squared deviation for every session length value. After summing up them, the result is divided by count of sessions, and square rooted to yield the standard deviation.