

Assignment 8 Writeup

There are two sections in this assignment.

Section 1 is a follow up of assignment 7. We modified the sequential processor in assignment 7 to do this part. The output files are five dat files:

- 1) liftRides.dat – contains detailed skier rides information, ordered by skier id. For the lift rides with the same skier id, order them by the time
- 2) skiers.dat – contains the skiers summary information, for each skier, stores the skier id, number of rides, total vertical, and number views
- 3) lifts.dat – contains the number of rides for each lift id ordered by the lift id
- 4) hours.dat – contains the top ten most popular lifts for each hour, records are ordered by hour id. For the same hour id, records are ordered by frequency of lift
- 5) *skierRow.dat – records the starting row index for the skier records in liftRides.dat (this file is to help read the records in liftRides.dat)

Section 2 will query the data generated by section 1, and save the results of the queries in 20 txt files. There are four types of queries, each corresponds to one dat file. The QueryProcessor creates 20 QueryThread to process the queries. Each thread is responsible for a disjoint segment of the queries. All threads start at the same time, and the termination is also synchronized.