Our plan to stage 8:

We started by extracting each user's different attributes like interarrival times, consumption, runtimes, think times and everything modeled in earlier stages.

Our next step would be fitting each attribute into a modal distribution like we've done in stage 6.

Then we can take these fitted distributions and by using the following syntax: "UserID 'attribute' 'distribution_name' 'param1' param2' param3'" we can push these lines to our configuration file.

After simple calculations, our log has about 80-100 different users so assuming for each user there will be a maximum number of 5 attributes then our input file will consist of a maximum 500 lines like the one described above (even 1800 is still okay because our log consists of 18239 jobs), and some additional input. (Like distributions of users' activity, residence times, and more). After that we can start to write our code using the User Resampling method described in chapter 8.3 of the book.

As a first step we decided to divide into 2 teams, first team that works on extracting appropriate data and fitting it into modal distributions and to prepare the configuration file, while the other team assumes the correctness of the configuration file and could starts coding to make the User Resampling.

As we remember in an earlier lecture was mentioned that to use user-based modeling the log should be at least 3 years old or something like that. While on the other hand, our trace is 92 days old. So, an alternative way we were thinking about is HMM (hidden Markov model)- maybe represent each state with the user group ID.