

CS 6240: Project Progress Report

Goals: (1) Practice an important job skill: presenting highlights of your work to others. (2) Re-evaluate project goals if necessary. (3) Make sure everybody is making good progress.


Each team has to create all deliverables from scratch. In particular, it is not allowed to copy another team's code or text and modify it. If you use publicly available code or text, you need to **cite the source** in your report!

All deliverables for this HW are due as stated on Canvas. This HW is worth 100 points and accounts for 15% of your overall homework score. Please submit your solution through Canvas and make sure your report is a **PDF** file. Do not submit anything else (source code, result files, log files). The progress report **cannot exceed 5 pages**

File naming convention: use the last names of all team members, sorted alphabetically, as the file name. For instance, the two-person team Joe Smith and Mary Miller would use file name MillerSmith.pdf. Both Team member needs to upload project progress report on canvas.

Project Progress Report

The progress report serves two main purposes. First, it gives you the opportunity to provide information about your progress. Second, you can use it as the basis for the final project report, saving you time later when you need it most. Make sure you include the following:

- Briefly describe the data and show a few lines of actual data in your report. You may copy this section from your proposal.
- Briefly describe each task you have completed so far or have spent a significant amount time on.
 - Show the pseudo-code for each of these tasks. Do not just copy and paste source code! (Exceptions are SQL, HiveQL, and PigLatin, which by design are concise and hence can be included directly if appropriate.)
 - Show concrete results, including running times of your programs. E.g., for a sampling task discuss properties of the sample. For an analysis task, show the corresponding results in graphs or tables. You want to show convincing evidence about your progress.
 - If you already performed a careful analysis of your program and maybe improved an initial program design based on this analysis, also include this discussion. Example discussion:
 -  We initially wrote program X. [Show pseudo-code and briefly explain program X.]

- ✎ Then we realized that it took much longer than expected. [Provide a few concrete numbers.] To find out how to improve X, we looked at the logs and realized that X transfers 10 times the input size from mappers to reducers.
- ✎ To reduce the cost of the shuffle-and-sort phase, we designed program Y. [Show pseudo-code and briefly explain program Y.] We believe it will address the problem because...
- ✎ We have not completed testing of program Y yet, but will include experiments comparing X and Y on our data in the final report.
- Briefly discuss any major problems you ran into, which required major adjustments to your original proposal. Make sure you clearly discuss changes you want to make compared to your initial project proposal.
- Briefly discuss the remaining tasks that you will work on until the end of the semester.

Deliverables

1. The project progress report as discussed above. (1 PDF file)