

Project Schedule

Each undergraduate student is expected to propose and implement one large MapReduce project using the test machine (for debugging on a lesser amount of data) and on AWS for a larger run. Here is an approximate schedule of what should happen, listed by *milestones* (see Milestone (project management)).

3 Apr

Selection of some Big Data -- please email your selection.

8 Apr

Proposal of project, what will be done with MapReduce to query or analyse the data. This should be at least a paragraph stating the goals.

10 Apr

Pseudo-code for a *fallback* (or minimal) attempt at the project. So if the grand plan fails for some reason, you have at least some minimum accomplishment. Also, extract some small amount of the Big Data. What is small? Typically this will be just a few gigabytes, not hundreds of gigabytes.

15 Apr

First mrjob of the minimal implementation on the test machine.

17 Apr

Running on AWS! Show that you can run your minimal implementation in the cloud.

22 Apr

Start in on the more ambitious implementation (deliver at least some code on this).

24 Apr

Running on test machine.

29 Apr

Running the big job in the cloud.

1 May

Polishing, documentation.

5 May

Demo the big job in class presentation.