P3 Handout

Revised March 4, 2013

Objectives

- To explore the practice of big data analytics
- To demo and present your work to the rest of the class in a poster presentation
- To produce a research report documenting your work (including analysis, design, and related work)

Project Ideas *(Each Project Must Receive Instructors' Pre-Approval)*

There is no one standard project in this course. Each team, preferably of size 3, will need to work with the instructors to develop a project. The project must build a solution for a big data analysis problem using a data-intensive handling framework (e.g., map-reduce, Hadoop, Pig or others) and then applying appropriate data mining techniques to solve the big data problem. Each team may make use of cloud services from vendors such as Amazon Web Services or Microsoft Azure to allow their solution to scale to magnitude needed to solve their big data problem. Each team will have a great deal of control to make decisions within these broad guidelines here, but working closely with the instructors throughout the quarter is recommended.

A well-designed and implemented project will yield results worth presenting at the RIT Annual Graduate Symposium. In addition, if a team really wants to, the work done on this project can be targeted for a real conference outside RIT. Under this option, the team would need to work closely with the instructors with more flexible deadlines; of course, the final submissions and poster presentation dates must be met. Make an appointment with the instructors if your team wants to pursue this option.

P3 Evaluation

Step 1: Team Score

Phase 0	5%	
Phase 1	20%	
Phase 2	25%	
Phase 3	25%	
Research Paper	25%	
Overall team score	**100%**	
Scaled team score	**70%**	(70% of overall team score)

Step 2: Individual Score

Team score to individual adjustment	± **a**	Adjustment up or down based on (a) peer evaluation feedback, and (b) individual P3-related performance during demo and quarter (as observed by instructor).
Individualized team score	70%	Scaled team score ± **a**
Team peer evaluation	*-20%*	Only if you fail to submit a well-detailed overall peer evaluation
Presentation/Demo	10%	
P3 questions during final exam	20%	If a student's performance on P3-related questions is significantly better or worse than her or his team's, individual grade will be further adjusted up or down.
Final P3 score	**100%**	See P3 component weight in the Course Syllabus handout.

Initial Schedule and Deliverables *(Submission dates on myCourses are more accurate!)*

Date	Event	High-level Description
March 12	Team formation	• Select your teammates, and a name for your team.
March 21	Phase 0	• Submit a 1-page description of your project including your sample cloud system and application. As the final report is to be written up as a research paper, you need to start working toward this goal by using the ACM style template linked on myCourses. This will allow you to revise the paper throughout the quarter for each phased deliverable.
April 11	Phase 1	• About 2-3 pages describing what has been done, what will be done, how it will be done, and the plan to cover topics stated above. Some details about your implementation (code, SQL, etc.) need to be included in the submission. • Submit the revised version of your research report, which should be a revision of your earlier submission. • Zip up any documentation, source code files, etc.
April 26	Phase 2	• Submit the essential parts of your system and application, along with documentation of your current accomplishments. • Submit a revised version of your Phase 1 report to reflect the progress made since the Phase 1 submission, and changes anticipated in the final submission.
May 06	Poster	• Soft-copy of poster.
May 07 (Last class)	Mini-conference	• Poster & demo. It will be organized like a poster presentation at a mini-research conference.
May 07	Research Paper	• Research paper must cover project motivation, architecture, design, implementation, lessons learned and future work, and should be written as an ACM conference paper.

May 07	Phase 3	• Final submission should include:
		  • Research report covering project motivation, architecture, design, implementation, lessons learned, current status, and future work, written up as a formal research paper.
		  • Working documented source code of system/application.
		  • A README file that describes how to set up your project submission, and also explicitly describes the main items of work done by each team member.

May 13	Overall peer evaluation	• Each team member will confidentially report on the overall contributions of all team members; see Surveys area. • If you do not fill this evaluation out carefully and in detail, your individual project grade will go down significantly.
Blog	Through-out the quarter	• Maintain a regular (several times spread over each week over the quarter) blog about your Distributed Systems activities: what you read, what you are working on, and so on. ***Separate grading component!***

Submission Information

For each of the three major submissions (Phase 1 to 3), place all relevant files to be submitted in folders named phase0, phase1, phase2, and phase3 respectively. Each submission must include all relevant files (including those modified from earlier project submissions). Include documentation/instructions on how to run your application.