CSC 560	Modern DBMS Architectures	Alexander Dekhtyar

Assignment 2: Project Proposal

Due date: Thursday, October 11, in-class

Assignment

This is a team assignment. Each team will work on putting together one set of deliverables. This also is the first part of your course project.

Proposal

This assignments marks the commencement of your team course project. Before working on the project for the rest of the quarter, you need to come up with the idea, present it to your peers, defend it, and then spend the remaining time this quarter being completely consumed by it. We start by coming up with the idea and presenting it.

Coming up with a project idea. Usually, a course project in a graduate level class needs to rise to the level of a *somewhat publishable (with some possible extra work after the quarter is over) piece of research.* This works really well in a course where there is one major project and all other assignments are short-term, do not involve programming or significant design.

In this course we will have a number of 2-3 week-long assignments, that would need to be done in parallel with the course assignment. Because of this, we nallow for a somewhat wider range of team course projects.

What kind of project? Each team can propose a project that meets one or more of the criteria below.

- **Empirical study.** Pick a research question from the field of databases. Formulate a hypothesis. Design an empirical study. Implement the study (this may involve software development/system building). Evaluate the results.

Example. Research question: which of two different XML shredding techniques (decomposing XML documents into relational tables) produces better results for XML join queries? Study: implement two shredding techniques, implement XML join queries on top of them, create/find a proper dataset and a benchmark, perform measurements.

- **System-building project.** Pick a DBMS design topic of interest and a challenge task. Build a proof-of-concept software system that implements the chosen task (meets the challenge). Perform a performance/functionality/feasibility analysis of your system and report on it.

  Example. Challenge task: implement timestamp-based distributed transaction management that preserves atomicity, isolation, durability and eventual consistency over sharded distributed databases. Project: create/adopt/adapt the transaction management algorithm(s), implement the algorithms. Build a testing framework that simulates the work of the remainder of the DBMS. Prove atomicity, isolation, durability and eventual consistency of your implementation.

- **Application-driven project.** Pick an application or application domain, *that presents a clear set of challenges for relational database modeling techniques*. Design an alternative solution for the the application/ application domain, that uses various other data management approaches found in class. Build a prototype, if you have time.

  Note: You cannot simply take a token "bank database" or "e-commerce site database" and design/build a database app. You must

  1. bring in a *real* application or application domain, with real data and (if possible) a real customer; the problem/problem class itself must have a significant amount and diversity of data;
  2. demonstrate (as part of the project) the problems with relational modeling of the data in the application/application domain;
  3. present a clear analysis of why your solution/solutions will improve upon the relational databases for this application/application domain/class of applications;
  4. either implement a proof-of-concept prototype, or provide a clear software design and implementation roadmap.

  Example. Application domain: integration of geolocation data with US Census data for on-line real-time mashups. Project: determine the bottlenecks of representing, storing and managing all data on a single-server RDBMS. Design an alternative data representation approach, build a data model reflecting this approach. Implement (or create a proper software design document for) the prototype of the data representation and storage with enough functionality to demonstrate how key use cases of this class of applications are handled.

Non-functional suggestions. Some tips and suggestions for picking a topic and fleshing out project ideas.

1. **Choose an interesting problem.** The key reason for you reading a bunch of self-assessment and prediction papers from the last 20 years is to get you to realize the variety of challenges the field of databases faces. The hope is that each team can find something it feels like an interesting task/challenge and go with it.

2. **See if this can be of real benefit to someone on your team.** You a team member has a data-management-related project ahead of him/her, allow this team member to "bring" it to your attention and consider if helping out on a task like this would be a worthy cause.

3. **Future work.** If you can, pick a topic that you would not mind working on outside of the class, if appropriate rewards present themselves. A typical reward is the ability to publish the work[1], but other rewards (a thesis topic, a startup idea, a way to get hired by Facebook or Oracle, etc.) are possible.

4. **Be realistic.** This is important. Don't take up things that require 4-year committments for the entire team. Figure out, how many person-hours you can devote to the project throughout the quarter, and use it as the balancing factor when discussing the scope of the project amongst yourselves.

5. **Talk to me.** I am happy to discuss your ideas with you as you develop them. If you need to run something by me in terms of whether this is an acceptable topic, scope, direction — talk to me *before* the assignment is due. This way you may be able to incorporate the first round of comments from me directly into your proposal.

What to Submit and When to Submit It

Proposal Abstract. By **Tuesday, October 9**, each team needs to submit its proposal abstract - a short, one-paragraph statement outlining the project idea/direction. I need something I can read and react to quickly.

Submission procedure: email dekhtyar@calpoly.edu, **and** cc: your entire team. The email must be in my mailbox by 11:00am on **October 9**. Also, bring a hardcopy of your abstract to class and give it to me. Put the tentative project title, names of all team members and the paragraph in the body of the email *as plain text*.

[1] Each previous CSC 560 resulted in at least one project that was taken further by its team. In one case, the end result was a short paper at a cool conference. In another case, the end result was the team's participation in the CSU Research Competition. While I wound up as the thesis advisor for two of the eight students involved in these project, the theses I advised had nothing to do with the CSC 560 work — the students committed to continuing the project simply because they liked what they were doing.

I will respond to your email by the end of the day Tuesday, October 9 with comments and suggestions about your proposed project topic. The response will be primarily based on the abstract, but might also incorporate any discussions I may have with the teams.

Proposal. On **Thursday, October 11**, submit your proposal documents. The proposal document must be formatted as a technical report paper (single column, single-spaced text), must have a title, a list of authors with affiliations and email addresses, an abstract (the one submitted, or an improved version), and a brief description of the proposed project. The description shall contain:

- **Introduction.** A brief introduction must be included. Outline the reasons for choosing the specific project, the problems/challenges the project will address, the results you hope to achieve.
- **Background/Related Work.** Provide a **brief** background on the subject area of your project and any other information that you might consider important. If discussion of related work is warranted, provide some examples of related work with brief explanations of percieved relevance (don't treat this as a complete related work section — rather as a "we believe this work is important, will tell you more in our project writeup" section).
- **Proposed project specifics.** Desribe briefly what you want to accomplish, and, if you already know it — how. Outline, how you will be evaluating the success of your project.

I prefer that you write your papers for the course in LaTeX, however, Word documents (and `.odt` documents) will be accepted as well. You

Submission procedure. Submit a PDF version of your proposal paper via email sent to dekhtyar@calpoly.edu by **8:00am, Thursday, October 11**. (I need to print the submissions before the class starts).

Presentation. We will spend one half of our **Thursday, October 11** class discussing the project proposals of each team. Each team should prepare a brief, 5-minute presentation discussing the proposed project. Each team gets a total of 10 minutes to present and answer followup questions. **Please, delegate AT MOST two people on your team to actually present**. I do not want all five-six members of the team participating in a 5-minute talk. All team members are welcome to participate in the followup Q&A.

Submission procedure. You are allowed to use slides. I would like the total number of computers used for presentation purposes kept to a minimum (or really efficient computer switches) - so having an on-line version of the presentation that can be downloaded to any laptop (or having a flash drive the the presentation slides) is a positive. Given the length (or, to be more precise the *unlength*) of the presentation, don't go overboard on slides. Try to make your slides clear and illustrative. You do not need to submit the slides ahead of time, but if you want to email them to me or to email me a link to an on-line version of the slides, before the class starts on Thursday, I won't mind.

Note on delivarables and grading. While your grades and my *written comments* for the proposal are between me and individual teams, the document you are submitting is considered public. It will be posted on the course web page, and students from other teams will be asked to read the proposals and provide some written feedback in the next assignment.

Good Luck!