Design Project
Stage 0

Overview

This document provides a brief outline for the course design project. Further documents will be released to provide you the problem description and to guide you through specific stages of the design.

The design project is to be performed by teams of four people each. Five teams will be formed. Team formation is left up to you. I recommend forming a team with an eye for completing both this project and the supplemental data analysis project, rather than having different teams on these projects.

The project will have the following timeline:

No.	Activity/Assignment	Dates/Due Date	Comment
1.	Initial discussion	*April 24, 2012*	brief discussion during the lab
2.	Team formation	*by April 26, 2012*	
3.	Problem description released	*April 24, 2012*	
4.	Stage 1 (preparation)	*April 24-May 1, 2012*	
5.	Discussion	*May 1, 2012*	*tentative*
6.	Stage 2 (design)	*May 1 - June 1, 2012*	
7.	Report due	*June 4, 2012*	
8.	Presentation	*June 5, 2012*	final exam slot

We may use a full lab or part of a lab session some time in May to conduct more discussions on the project.

The Problem: Smart House Monitoring

Scenario.

Utility companies have been measuring the consumption of resources (electricity, gas, water) for individual households for years and years via the means of meters that measure the intake of each type of the resource.

However, technology is around the corner to allow for monitoring of the resource consumption for ***each individual appliance*** in a house. This forms the basis of our scenario.

In what follows, we track the *appliance-by-appliance* use of the three key resources (utilities) provided to individual households: electicity, water and gas. We **will not be monitoring** any specific communications (e.g., the content that comes via phone lines, television cable or satellite broadcast). We also will not consider *sewage* use. We use the term *appliance* to describe any household device that is capable of consuming one or more of the abovementioned resources *independently* of other devices.

For example, a microwave oven, a television set and a computer are appliances that consume electricity. A kitchen faucet and a bathtub are appliances that consume water. A stovetop, an oven and a dryer are appliances that consume two types of resources: gas and electricity.

In every household in a city, all appliances that consume water, gas and/or electricity are fitted with a small metering sensor that detects the resource consumption. For appliances consuming multiple types of resources, consumption of each time (e.g., of gas and electricity for a dryer) is metered independently.

Every sensor is equipped with a short-range bluetooth transmitter, which connects to a blutooth and 3G/4G-capable transmitter located on the main meter for the house (there is one main meter for each resource per house). From the main transmitter, the information gets submitted to the servers of the appropriate utility company.

Software to Design

The information collected from each household resides on the servers of the utility companies (power, gas, water companies). The new regulations require this information to be shared via a data clearinghouse, which can import any resource consumption data collected from the system described above. Additionally, the data clearinghouse can collect other household related data available from a variety of sources (e.g., property records, geolocation information, etc...).

You are the software design team of the data clearinghouse. Your task is to design the data clearinghouse system that collects data, analyzes it and provides important and useful services to its customers.

The role of KDD

At the heart of your system is a vast collection of highly detailed, and extremely important and sensitive data. This data can be used for a wide range of tasks, related to monitoring of resource consumption (and thus potentially making the planet greener) and other tasks.

The collected data affords a **large array of possible uses** and **it is the goal of each team to determine which data will be used and how**.

Legal Issues

Resource consumption data collected via this system is very sensitive, as it can provide a very detailed description of what is happening inside a specific household at each specific moment of time. As such, the use of this data is highly regulated in the USA, and specific detailed policies exist. Your software system must abide by the laws, rules and regulations, guiding the use such data in the USA.

It is each team's responsibility to become acquainted with the appropriate laws and regulations.

Note: Each household whose resource consumption is monitored in the way described above is considered to be one category of users. Your system shall come with its own **End-User License Agreement** (EULA). Some data use issues may be stipulated in it (notice that you may need other EULAs for other types of users of the data clearinghouse software). While no team needs to write a full EULA for this project, it might be a good idea for each team to decide, which data use policies must be covered by their EULA.

The Assignment

Stage 0. On the preliminary stage of the project, the following tasks shall be performed:

1. **Team Formation.** Teams shall be formed as soon as the assignment is released.
2. **Stakeholder determination.** Each team shall start the design of the project by coming up with a list of stakeholders for the software system. For the purposes of this assignment, a *system stakeholder* is a category of system users, whose interests will be coded as use cases for the system (and, eventually, implemented).
3. **Services determination.** For the selected stakeholders, each team needs to prepare a preliminary list of services the data clearinghouse system will be providing.

4. **Dataset determination.** Each team shall determine what information will be collected and stored in the data clearinghouse.

On Stage 0, your form teams. Teams are expected to contain 4-6 people each. We cannot have more than 9 teams in the class (due to the final presentation requirements and time alotted). It is preferable to have even fewer teams (each team gets more time for presentation). Stage 0 deadline is **Thursday, April 26.** Each team needs to prepare the following deliverables:

1. **Team wiki page.** The wiki page shall, at the very least, contain the team name, list of team members, and (eventually) links to all other required deliverables for all stages.

On Stage 1 (April 26 -May 1), each team shall conduct preliminary meetings and discuss how it wants to proceed with the system design. The deliverables for this stage are:

1. **An initial outline of the project.** A short document briefly outlining the direction in which the team plans to take the project. In particular, the document must specify
   - The stakeholders chosen for the project.
   - The interests/conflicts/desires of individual stakeholders.
   - The services to be provided by the software system.
   - The data you are planning to collect from the system.
   - The analytical tasks you are planning to perform.
2. **Initial Presentation.** Each team shall prepare an 7-10-minute presentation based on the initial outline of the project. The presentations shall be accompanied by some visuals (slides), which should also be posted to the teams' wiki pages.

We will conduct a discussion session on Tuesday, May 1 or on Thursday, May 3, in which each team shall present their initial plan, and other teams will have a chance to ask questions and discuss the project.

Stage 2. Full stage 2 assignment will be released upon completion of Stage 1. On Stage 2 of the project, the teams will finalize the design of their systems and will prepare two major deliverables:

1. **Final design document.** The actual software system design, clarifying, expanding on and extending your initial outline. The exact specifications for the final design document will be made available to you around **May 3**. **The final design document is due Monday, June 4.**

2. **Final presentation.** Each team will give a 15-20 min. presentation of their design in class during the finals week. The exam time is **Tuesday, June 5 7:10am.** We will use this time for presentations and followup discussion. The presentations are due **June 5.**.