

CSE 241
Database Systems
Spring 2005
Project 3

Distributed: Tuesday Feb 15, 2005

Checkpoint 1: Tuesday Mar 1, 2005 (E-R design review)

Checkpoint 2: Tuesday Mar 29, 2005 (relational design review)

Project due date: Tuesday Apr 12, 2005

Please submit your work both on paper in class and the online parts by email to sal7 with cc to your instructor. Be sure the email and paper versions correspond.

Project Overview:

The goal of the project is provide realistic experience in the *conceptual design*, *logical design*, *implementation*, *operation*, and *maintenance* of a small relational database using the same PostgreSQL system you used earlier.

Application Description:

The application consists of the operations of a real-estate office. The office needs to keep track of agents, buyers, sellers, properties on the market, and recently sold properties. This office focuses on homes rather than business real estate. The management of this real-estate office is not very computer literate. You are being asked to design the database, populate it with sample data (the management won't allow you to test with live data because of privacy concerns), and to write several SQL queries to demonstrate the system. The real-estate office is soliciting approximately 50 proposals, which it will evaluate starting April 12, 2005.

To learn more about the application domain look at some real-estate web sites. We're not pointing you to a specific one (lest we overload some small agent with a 8MB Pentium II) but the site

- <http://www.lehighvalleyvpa.org/default.aspx?pageid=8>

has links to several local offices that may help you. Please do NOT email these organizations. Just reverse engineer the database requirements from the data presented on the web sites. The queries we list below provide some useful hints as well.

Project Requirements

This section lists the things that you need to do. See below for a specific list of what to turn in.

1. Entity-Relationship Model
 - Construct an Entity-Relationship model representing the *conceptual* design of the database. At minimum include all of the entity sets mentioned in this handout, include a primary key for each entity set plus at least those attributes mentioned in the handout, and all appropriate relationship sets. You may make your design more complete based on your study of real-estate web sites.
 - Include primary keys in the E-R diagram, and show the cardinalities of the relationships.
2. Relational Model

- Convert the E-R schema to a *relational database schema*. Be sure to include appropriate indices and constraints. Be sure your relational schema conforms to your E-R schema. If you change your design later in the project don't forget to update the E-R schema.
3. Create each of the tables in SQL in your own database within PostgreSQL
 4. Populate your tables with a "sufficient" number of data items. HINTS:
 - Look ahead to the queries and choose data such that each query will have an "interesting" answer (e.g. not the empty set). Generate your own synthetic data set rather than trying to get data off the web. We don't want any nasty phone calls from realtors! You may find it convenient to write a program in C++ or Java to generate the data. The bulk load features of PostgreSQL, such as the copy command <http://www.postgresql.org/docs/7.4/static/sql-copy.html> may be of help here.
 - In order to keep the work level manageable, create images for only 2 or 3 houses (including the most expensive one) and let the image be null for most houses. You'll need blobs for this and the lo_import and lo_export functions which are described on the web at the following URL: (http://www.postgresql.org/files/documentation/books/aw_pgsql/node96.html). The images need not be of real houses on the market!
 5. Queries:

Construct and execute SQL queries for the following:

- a) Find addresses of homes for sale in the city "Bethlehem" costing between \$200,000 and \$250,000.
- b) Find addresses of homes for sale in the school district "Parkland" with 4 or more bedrooms and no swimming pool.
- c) Find the name of the agent who has sold the most property in the year 2004 by total dollar value.
- d) For each agent, compute the average selling price of properties sold in 2004, and the average time the property was on the market. Note that this suggests use of date attributes in your design.
- e) Show a picture of the most expensive house(s) in the database.
- f) Record the sale of a property that had been listed as being available. This entails storing the sales price, the buyer, the selling agent, the buyer's agent (if any), and the date.
- g) Add a new agent to the database.

Add additional queries to cover interesting data that you captured that we did not consider above.

What to turn in (both on paper and via email unless stated otherwise):

1. Your E-R diagram (on paper only – no need to create an electronic document, though you are free to do so).
2. Your relational schema. Either print out the **create table** SQL statements you used or use the \d feature to print out this information.
3. Do NOT submit the contents of all your tables. Just run **select count(*) from r**, for each relation r in your schema so we get a sense of how big your tables are.
4. Submit the queries in the format
 - Query in English
 - SQL query
 - answer

