

**Name:** \_\_\_\_\_

**Overview:** This assignment asks you to build a database-backed Web site based on the simple application you built in the last assignment. The goal is to gain experience in setting up a Web server and programming the interfaces of the Web server, your application code, and the DBMS (and any other modules you may decide to use). The Web application in this assignment is similar *in spirit* to those on many popular Web sites (e.g., amazon.com, ebay.com); however, several important issues, notably scale and efficiency, are ignored.

Functionally, the Web application in this assignment is very similar to that in the previous one. The Web interface must allow users to invoke all the functions in that assignment (e.g., connect, createTables, addBox, searchProductsByPhone) and view their results. The semantics of the functions remain unchanged. The main difference is that user interaction will use HTTP and HTML instead of the stdin-stdout interface of the previous assignment.

**Software:** By default, we will use *PostgreSQL* as the DBMS, *Apache httpd* as the Web server, and *CGI* as the interface to the Web server, all on the *linuxlab* cluster. You may use a programming language of your choice. As always, you are free to use any libraries and code you find helpful, but all code you use must be *clearly and prominently* acknowledged in your submission (both in the packaged code you submit and on the Web pages of your application produced using your submission). As always, no matter what resources you use, you must be able to explain how everything works. (“I call function foo with these parameters and library bar takes care of everything” is not a satisfactory answer.) If you prefer using different (or additional) software, you are free to do so. (For example, students in previous semesters have used Tomcat, PHP, JSP, and a host of other software.) The only catch is that you will be responsible for making sure all the software you use works with the submission procedure below, implying your Makefile must result in the installation and configuration of any extra software you need. Unless you are familiar with Apache and Web programming in general, **please budget enough time to learn about them**. Learning these interfaces is the major part of the assignment; the rest is essentially PHW02.

**Web Interface:** The entry point to your application is a page we shall call the *application home page*. It should include the following:

- **Identifying information:** A title (CMSC424-0201 Fall 2004 PHW03) and your name, preferred email address, detective cluster account, Oracle account, linuxlab account, and PostgreSQL account. (Do **not** include sensitive information such as passwords, student-ID, and SSN.) The title and your name should anchor links to the class Web page and your personal Web page, respectively. (The contents of your personal Web

page are irrelevant for grading as long as it is clearly your own page. If you don't have one, now is a good time to create it.)

- **Function links:** For each function, there must be a link that points to its *function page* (described below). The link's anchor-text must be the name of the function. The function links must be in the order the functions are listed in PHW02.

All the HTML pages used by your application (e.g., the function pages described below) must include a **link to the application home page** somewhere near the top of the page.

**Function Pages:** As mentioned above, there is one function page corresponding to each of the user functions of PHW02. Each function page should include the following:

- **Identifying information:** This information consists of all the identifying information on the application home page followed by the name of the function (e.g., *searchBoxesByLoad*).
- **Input fields for function arguments:** For each of the arguments the function takes, there should be one text input box. The boxes should be stacked vertically and should include descriptive labels. For example, use labels such as *Minimum Load (pounds)* and *Maximum Load (pounds)* for the arguments of the *searchBoxesByLoad* function.
- **Proceed button:** The above should be followed by a form-submission button labeled *Proceed*. When the user clicks this button, the appropriate function should be invoked with the user-provided arguments (e.g., *searchBoxesByLoad* with the minimum and maximum load values in the input boxes of the form). The result should be presented as a *result page*, described next.

**Result Pages:** These pages are used to display the result (output) of a function invocation. Each result page should include the following:

- **Identifying information:** This information consists of all the identifying information on the function page followed by the arguments used in the function invocation. For example, for the function *searchBoxesByLoad* invoked with arguments 10 and 20, the page should indicate *searchBoxesByLoad(10, 20)*. You are free to use an alternate formatting of this information (e.g., HTML tables), especially for longer, string-valued arguments. However, the information should be easy to discern.
- **Database Results:** The output for each function should be in a tabular format reminiscent of the output format described in PHW02. Instead of using separator strings (space-colon-space; see PHW02) and newlines as column and record separators, you should use HTML tables with descriptive column headings. For example, the result page for the *searchBoxesByLoad* function should contain an HTML table with two columns (and as many rows as there are result tuples). The first column should list the matching loads and the second should list the box's name; suitable

headings are *Load* and *Name*, respectively. The output should be sorted as described in PHW02. Further, wherever a PID is listed (e.g., in the second column in the output of *searchProductByDesc*), that PID should **anchor a link** leading to the details of the corresponding product. (Thus, clicking on a PID should result in the invocation of the *detailProduct* function with that PID as argument, in turn resulting in the display of the corresponding results page.

- **Diagnostics:** An area below the table should be reserved for diagnostic messages. You are encouraged to catch as many errors as possible. At the very least, you must catch errors arising from missing inputs, invalid inputs (e.g., a non-number in a field that requires numeric input), failed authentication (invalid username or password), and server unavailability. You must also produce an informative message when there is no data matching the search conditions (e.g., “there are no boxes with load rating in the given range” when no boxes match an invocation of *searchBoxesByLoad*). While you are not required to provide thorough error recovery in this assignment (meaning you can earn the full score by catching only the above errors), you are strongly encouraged to do so because you will need something similar when you implement your team project.

**HTML Design and Validation:** You are encouraged to create interesting designs for your Web pages subject to the above constraints. However, please make sure that all the relevant information is easily legible. Do *not* use (client-side) active features such as Java applets and Javascript. Further, all HTML should be validated using the W3C service at <http://validator.w3.org/>. (You may substitute another validation service if you prefer it; however, you must ensure that it is at least as strict as the W3C one.) At the bottom of each HTML page generated by your application, there must be a link labeled *validate*. Following that link should present the output generated by the W3C validator with that page as input. Obviously, you should check that this output always indicates that no errors were found. Make sure you get rid of all errors and try to get rid of all the warnings. You will lose points for errors, but not for warnings (although you should try to get rid of the warnings too). For an example of how this process should work, see the link labeled Check near the bottom of the class Web page.

Creating the validate link as described above may be harder than it first appears since many of the pages will be dynamically generated. Depending on how you set up your Web application, it may be difficult to run the validator by using the URL used by the live application. You may use a workaround such as storing the generated page in a temporary file and passing the validator a URL for that file.

**Submission:** The electronic submission procedure is similar to that used for PHW02. There is no hardcopy submission; we will grade the most recent file uploaded by each student. Electronic submission is by anonymous FTP to `ftp.cs.umd.edu`, directory `/incoming/cmssc424-0201`. You must submit a *gzipped tar* file containing the source files (*not* object files, class files, or machine code) required to compile and run your program and

any other software you use (other than Apache and other software that is installed on the linuxlab cluster). You should name this file `phw03-foo-bar.tar.gz`, replacing `foo` with your last name suffixed with initials (e.g., `SmithJK`) and `bar` with a 4-digit integer (e.g., `1984`). Unzipping and untarring `phw03-foo-bar.tar.gz` should result in the creation of a single directory (in the current working directory) called `phw03-foo-bar` (e.g., `phw03-SmithJK-1984`). Typing `make` in the `phw03-foo-bar` directory should result in the complete compilation of your program, producing two final executable files (machine code, shell script, Perl script, etc.): `startserver` and `stopserver`. Obviously, you will need to include a Makefile. You should also include a short README file describing the files in your submission; it should also include instructions for compiling and running your program. Please test very carefully that this unpacking and compilation procedure works with your submission. Your score will suffer greatly if it does not, or if your submission contains object files or machine code.

The program `startserver` will be invoked with a port number (e.g., `8424`) as the single command-line argument. This action should result in everything required to set up and start your application's Web server. The output of `startserver` should be a single line containing the URL of your application (e.g., `http://holmes.umd.edu:8424/`). Note that your program may be tested on any of the detective cluster machines (not just Holmes). The program `stopserver` will be invoked with no arguments. It must result in the termination of all the HTTP server processes used by your application. Be careful not to assume that your application will be the only one running on the test machine. If the port number supplied to `startserver` is unavailable, it should print out an appropriate error message and quit gracefully. (A better strategy would be for `startserver` to find and use an available port in the vicinity of the given one; you are encouraged, but not required to implement this feature for this assignment.)

**Cautionary note on running HTTP servers:** Please be very, very careful about how you run Apache (or another HTTP server) when you are debugging and testing your program. It is very easy to overload a machine with runaway `httpd` processes. Make sure you read, understand, edit, and use a `httpd.conf` file that does not overload the system. Pay special attention to the `MinSpareServers` and `MaxSpareServers` directives. Make sure you understand how to shut down the `httpd` processes by killing the proper (parent) process. Check your processes frequently using `ps -u yourusername`, especially before you log out (when you should kill all remaining `httpd` processes). You should read the manual pages for at least the `ps`, `top`, and `kill` commands. You should be very careful in this regard, mainly to be nice to other users but, as additional motivation, if you are found to be careless, you will lose points.

**Budget enough time for learning:** Most students find this assignment a *lot* harder than it first appears. *Many of the difficulties will not become apparent until you start working.* Please start very early and expect to learn a lot of new material. If you are not very familiar with Unix software and conventions (e.g., if you had trouble with environment variables or shell scripts), you should budget even more time.