CDF/DOC/SCMB/CDFR/3415

IMPLEMENTATION OF SMALL DATABASES UNDER UNIX

Liz Buckley-Geer and Glenn Cooper
November 8 1995

This note describes the setup that we propose for the CDF NOTES, CDF TAPES, and CDF PEOPLE databases under Unix. The goals are to meet all the requirements described in the requirements document for these databases (CDF Note #xxxx), to add some additional options, and to make access to the databases easier.

Location

Since the VMS machines will disappear eventually, the databases should reside initially on cdfsga, the central Unix machine for CDF.

Database product

We have developed test applications using mini-SQL (mSQL), a "lightweight database engine" that is optimized for fast, simple queries. mSQL is free for educational and research use. It supports a subset of standard SQL commands, with only a few extensions to the ANSI standard, so that applications written for mSQL can be easily adapted to a commercial or other database engine if needed.

mSQL may be accessed through its own terminal monitor program, through a C language API, or through interfaces to scripting languages including perl, tcl, and others. It supports client-server exchanges either through a Unix socket on the local machine or through a TCP socket from remote machines. Therefore, a user can access each database directly from any machine that has the mSQL software installed--something that is not possible with the existing DATATRIEVE package. (However, there are no VMS clients for mSQL, so access from VMS machines will be limited to the WWW and command-line interfaces described below.)

The fixed target experiment E831 at Fermilab has chosen mSQL for its calibration database, and reports that mSQL is reliable and meets the experiment's needs well.

User interface

Because of its combination of ease of use and machine independence, we propose using the World-Wide Web (WWW) as the primary access medium for these databases. (This will be in the protected part of the CDF area, so that only CDF members have access.) Queries to find, add, or modify records are easy to send as HTML forms to cdfsga (the WWW server), which can then perform the database queries and return the results through the Web. Users see the same interface regardless of the machine or operating system they are using. Those with terminals that support only ASCII text, or who have slow network connections, can use the Lynx Web browser, which shows an approximation to the standard interface.

For the tapes database, we will also port the current command-line (menu-based) interface to run under Unix, and modify the existing VMS interface. This will permit tape allocation by batch jobs and automated scripts, and also will allow a user who wants only to request a tape to do so without having to start a Web browser. The VMS command-line interface will allow simple requests for tapes and some level of searching, but may not support the full functionality of the WWW or Unix interface.

Advantages

Besides meeting the requirement to move to Unix, these choices of database engine and user interface offer several advantages over the current setup:

1. Users can access the databases from any combination of hardware and software, from any site, and will see the same interface, unaffected by the operating system of the database host.

2. The Web interface makes it easy to display defaults, lists of choices (e.g., group names for CDF Notes), and the nature of each entry: free-form text, yes/no, choice from a list of possibilities, etc. When modifying an entry, a user with a windowing terminal can use point-and-click editing tools, rather than having to retype the entry.

3. Mini-SQL is optimized for the kind of simple queries used for these databases, and searches are significantly faster than DATATRIEVE searches.

Test implementations

We have developed test versions of WWW interfaces for the RK (user) tapes and CDF Notes databases. These are accessible through the CDF VMS Migration page, at URL

http://www-cdf.fnal.gov/offline/vms_migration/vms_migration.html

By filling in blanks on a form, a user can search the list of tape assignments or CDF Notes; request a new tape or Note number; or modify an existing tape or Note record. We plan to make similar interfaces for other tapes databases and for the CDF People database.

The CDF Notes format has been altered slightly, combining the three Author fields of the existing format into one, and combining the two Pub_info fields into one. We have also added fields to indicate the revision number and revision date of the note.

Tests to date indicate that the mSQL back end is fast and reliable, and the interface seems intuitive and easy to use.

Making the switch

Moving the actual implementation of each database from the VMS cluster to cdfsga requires making a text copy of the database; copying this file to cdfsga; filtering it through a routine that writes each database entry as an SQL insert command; and installing the database under mSQL. Once the database is successfully installed and tested, the access routines can be changed either to a message pointing users to the appropriate Web page, or to revised versions that send rsh commands to cdfsga, as appropriate.

To accomplish these steps as reliably and with as little inconvenience to users as possible, we propose these steps:

a) A week before the switch, post a notice to cdfnews that the database will be unavailable for part or all of the day of the switch. Post a reminder the day before.

b) On that day, disable access to the database. Go through the steps described above to install the current version of the database on cdfsga.

c) If the old interface is to be duplicated, install on relevant machines the new version of the interface software that accesses the database on cdfsga, including a reminder that database access is now available through the Web. If not, replace the old interface with a message giving the URL for Web access, and a pointer to instructions for starting Netscape/Mosaic/Lynx.

d) Test access to the database. When all tests are successful, enable access and post a notice that the database is available.

CDF$PUB Access

For CDF Notes, we propose eventually moving the CDF$PUB area from FNALD to cdfsga. (This would be a separate step, after the databases have been moved). Posting will then be done from cdfsga through a mechanism similar to that used on FNALD, and will also be available through a Web interface. Posting will update the filename and date filed fields in the database automatically. The disk containing the files will be available on both cdfsga and FNALD.