

Gestión de bases de datos forestales

Jenny Salgado¹

Resumen

Se describe la importancia del diseño, gestión y los componentes de bases de datos forestales; los beneficios del manejo de información a través de bases de datos (acceso de los datos en tiempo real y por múltiples usuarios, disminuir la redundancia, integración de los datos, seguridad, sostenibilidad en el tiempo y el desarrollo de aplicaciones); los requerimientos y los componentes (lenguajes, metadatos, mecanismos de seguridad e integridad; factor humano) y finalmente, se indica la importancia de una política de acceso y uso.

Palabras claves: Investigación forestal; bases de datos; gestión de la información; almacenamiento de información; difusión de información; acceso a la información; políticas.

Abstract

Management of forest databases. This describes the importance of design, management and database components in forest research; the benefits of information management via databases (access to data in real time and by multiple users, decreased redundancy, integration of data, security, sustainability in time and development of applications); the requirements and the components (languages, metadata, security mechanisms and integrity; human factor), and finally, it points out the importance of a policy for access and use.

Keywords: Forestry research; databases; information management; storage; dissemination of information; access to information; policies.

¹ Proyecto CATIE - Finnfor. jsalgado@catie.ac.cr

Introducción

En la investigación forestal, el uso de bases de datos ha venido tomando importancia debido a las ventajas que ofrece el mantener la información ordenada; entre ellas, que múltiples usuarios puedan hacer consultas en tiempo real y/o de manera simultánea mediante criterios previamente definidos.

Por lo general, los investigadores han almacenado la información en cientos o hasta miles de archivos de hojas de cálculo (Calc de OpenOffice, Excel de Microsoft Office...). Si bien este sistema ha sido de uso común durante muchos años, ya que permite ingresar, mantener y trabajar con todo tipo de datos (texto, numéricos, alfanuméricos y fórmulas), su principal limitación es el manejo de grandes cantidades de datos. Esta limitación se vuelve un serio impedimento cuando se trata de proyectos grandes y complejos, o de investigaciones a largo plazo. Asimismo, una colección de archivos no es una base de datos, toda vez que, por definición, la BD requiere de un sistema de administración de los datos.

Actualmente existen programas de administración de datos con licencia libre, tales como MySQL, DBF Viewer 2000, phpMyAdmin y PostgreSQL 8.3.3. Estos programas permiten no solo almacenar información sino también ordenar los procesos de diseño, actualización, consulta y mantenimiento de la información en la BD. Estos procesos son los más importantes y frecuentes en la administración de una base de datos.

La creación de una base de datos se inicia con el diseño; para esta etapa del proceso se deben considerar todos los datos disponibles de una investigación. Con frecuencia, el diseño se convierte en un serio reto, ya que por lo general hay que realizar pasos adicionales de comprobación de los datos y/o completar

información perdida, inexistente o inaccesible. En la página en internet de MySQL se ofrece información detallada acerca de cómo crear y utilizar una BD (<http://dev.mysql.com/doc/refman/5.0/es/database-use.html>).

La administración de la información mediante un sistema apropiado de bases de datos ofrece ventajas significativas, ya que en tan solo unos segundos –y mediante criterios de búsqueda previamente definidos se puede recuperar información antes ingresada, buscar y encontrar experimentos particulares, o metadatos y variables. Este manejo ágil de la información permite hacer análisis puntuales de los datos obtenidos.

Bases de datos relacionales

Por principio, toda investigación busca coleccionar y/o generar datos. Tradicionalmente, la forma más práctica de recopilar y almacenar la información han sido las tablas y cuadros de datos creados mediante hojas de cálculo que se almacenan en archivos independientes. Si bien el conjunto de archivos constituye un banco de datos, no es una base de datos relacional. Para conformar una base de datos relacional es necesario que los datos que pertenecen a un mismo contexto estén referenciados entre sí (Sánchez 2004). Es decir, una base de datos relacional es una colección de registros, archivos y otros objetos que son almacenados y organizados de tal manera que se reconozcan sus interrelaciones para facilitar el acceso.

Cada BD está compuesta por registros subdivididos en campos. En el diseño de la base de datos es necesario definir de previo los registros y los campos que la van a conformar. En esta etapa es importante considerar los tipos de datos que se van a ingresar en cada campo. Existen muchos tipos de datos, aunque los más comunes son numérico, carácter, fecha y lógico.

Desde la etapa de diseño es también importante que se definan las tablas de referencia, tablas independientes y las interrelaciones entre tablas. El administrador debe conocer muy bien la estructura de la BD para garantizar el buen manejo y mantenimiento. Los sistemas actuales de bases de datos poseen distintas herramientas que permiten ver, de manera gráfica, el diseño de la BD, las relaciones entre tablas, las llaves primarias y foráneas y, en general, la estructura completa de la BD.

Por medio de archivos independientes es posible detectar algunos problemas, tales como aumento en la redundancia, inconsistencia de datos, aislamiento de datos, seguridad. Se recomienda utilizar sistemas de bases de datos para minimizar o evitar estos problemas.

Beneficios que aporta un sistema de base de datos

En años recientes, al mejorar el acceso a internet y con el desarrollo de aplicaciones, ha cobrado relevancia el uso de BD, en comparación con la administración de la información en archivos independientes. Quiroz (2003) y Codd (1970) destacan los siguientes como los principales beneficios que se obtienen con el manejo de información por medio de BD.

- **Acceso:** una BD permite el acceso en tiempo real y por múltiples usuarios; a través de internet se puede ingresar desde cualquier parte del mundo. Esta característica es muy favorable en el caso de organizaciones e investigadores que requieren acceder a la información al mismo tiempo y fuera de la oficina.
- **Disminución de la redundancia:** la información almacenada en archivos independientes puede albergar redundancias entre los datos; es decir, que un mismo dato esté almacenado en dos o más lugares, lo que causa inconvenientes a la hora de actualizar los datos,

e inconsistencias en los análisis. Además, se desperdicia espacio de almacenamiento.

- **Integración de los datos:** cuando los datos se mantienen dentro de una estructura ordenada, como una BD, las consultas toman menos tiempo y recursos, ya que se basan en criterios de búsqueda previamente determinados. En un proceso de recuperación de datos de diversas fuentes, la búsqueda se hace en una fuente a la vez y luego hay que unificarlas, con el agravante de que no siempre se tiene información compatible y/o comparable.
- **Seguridad:** con el fin de asegurar la integridad de los datos, los usuarios de una BD disfrutan de privilegios diferenciados, pero también de restricciones de edición o eliminación.
- **Duración:** la información almacenada en una BD puede ser reconocida por nuevas versiones del sistema de administración de bases de datos; esto asegura la sostenibilidad en el tiempo.
- **Desarrollo de aplicaciones:** mediante el acceso a los datos almacenados se pueden crear aplicaciones que ayudan a visualizar, modificar o manipular la información.

Requerimientos para establecer una base de datos

Para que un sistema de bases de datos pueda ejecutarse correctamente es necesario tomar en cuenta los siguientes requerimientos:

- **Tamaño:** los sistemas de bases de datos utilizan mucho espacio en el disco duro y también en la memoria principal (ram) de la unidad de cómputo.
- **Personal calificado:** para entender las capacidades y limitaciones de la BD, el administrador debe tener formación en programación.
- **Costo:** generalmente las licencias de *software* de gestión de datos son costosas. Afortunadamente,

ahora es posible tener acceso a programas libres que no generan costos fijos.

- **Instalación de equipo:** al adquirir una licencia de un programa para administrar una BD, también es importante adquirir equipo adicional como servidores, memorias, discos duros, etc.
- **Rentabilidad:** debido a la inversión necesaria en equipos, capacitación del personal y tiempo, la BD tarda cierto tiempo en estar operativa.

Sistema de administración de bases de datos

Según Ramos *et al.* (2006), los sistemas de administración de bases de datos permiten a los usuarios acceder y manipular la información; a los administradores, les ofrecen las herramientas necesarias para ejecutar tareas de mantenimiento y administración de los datos.

Las principales funciones de un administrador de datos, según Ramos *et al.* (2006), son las siguientes:

- Diseñar la BD; es decir, cómo se va a almacenar y organizar la información.
- Crear la BD y verificar el almacenamiento de datos en la base de datos.
- Recuperar datos mediante consultas y reportes.
- Actualizar los datos cuando sea necesario y cambiar los contenidos de la BD.
- Programar aplicaciones que mejoren la consulta a los datos por parte de los usuarios.
- Controlar la integridad de la BD.
- Monitorear el comportamiento de la BD para asegurar el acceso a los usuarios.

En la actualidad existen muchos programas y lenguajes de programación de bases de datos; entre ellos, MySQL, SQL Server, PostgreSQL, Access, Oracle, Fox Pro. Estos productos difieren en precio, rendimiento, facilidad de administración de bases de datos, accesibilidad y fun-

cionalidad; al seleccionar el *software* más adecuado se deben considerar todas las particularidades.

Componentes de un sistema de bases de datos

En prácticamente cualquier BD actual interactúan cuatro elementos esenciales:

- **Lenguajes:** si bien existen diversos lenguajes de programación de bases de datos, el más común es el lenguaje de consulta estructurado (Structured query language SQL). Este lenguaje permite especificar las operaciones que se pueden realizar entre las tablas relacionadas mediante:

- El lenguaje de definición de datos (*Data definition language* DDL) permite crear la estructura de la BD, considerando todos los objetos que se pueden incluir. Mediante los comandos CREATE, DROP y ALTER es posible definir la estructura de la BD.
- El lenguaje de manipulación de datos (*Data manipulation language* –DML) permite consultar y manipular la información depositada en la BD. A este lenguaje pertenecen los comandos SELECT, INSERT, UPDATE y DELETE.
- El lenguaje de control de datos (*Data Control Language* –DCL) permite asignar privilegios a los usuarios, confirmar o abortar transacciones. Estas operaciones responden a los comandos GRANT y REVOKE.

- **Metadatos:** este es el diccionario que contiene toda la información acerca de los datos; por ejemplo, la definición de todos los objetos de la BD, las columnas de cada tabla, vistas, procedimientos, ubicación física de los objetos y espacio asignado. También es posible incluir los privilegios, restricciones y roles asignados a los usuarios.
- **Mecanismos de seguridad e integridad:** por medio de este elemento se

asegura la ejecución de copias de seguridad de los datos, la protección de los datos ante accesos no autorizados y la recuperación de datos en caso de error del sistema, o cualquier otra situación imprevista.

- **Factor humano:** este elemento permite definir diferentes roles:
 - Usuarios finales que pueden tener acceso a la información sobre la cual tengan privilegios.
 - Programadores que crean aplicaciones con los objetos de la BD para los usuarios finales.
 - Administradores que garantizan el correcto funcionamiento de la BD y gestionan todos sus recursos. Tienen el nivel más alto de privilegios y responsabilidades legales en caso de que los datos estén legalmente protegidos.

Política de acceso a datos contenidos en una BD

En la gestión de una base de datos es necesario proteger y controlar el uso de los datos contenidos para evitar pérdidas, uso indebido o alteración

de la información. El uso de información científica debe estar ligado a una política de acceso y uso de datos, la cual permita al administrador controlar el buen uso de la información y garantice la seguridad de los datos.

El principal objetivo de la investigación científica es la divulgación y promoción de la información que se genera por medio de los estudios y experimentos establecidos. En años recientes se han generado políticas que norman el acceso y uso de los datos, siempre buscando un equilibrio apropiado entre el reconocimiento de la autoría de esfuerzos de investigación, por un lado, y el acceso y el uso de los datos, por otro.

En general, las políticas de acceso y uso de los datos toman en cuenta los aspectos siguientes:

- **Definición de los datos:** se especifica el tipo de datos que se incluyen en la base de datos; por ejemplo, datos económicos, sociales, ecológicos, etc.
- **Definición de la propiedad de los datos:** se detalla los tipos de usuarios,

permisos y restricciones para el uso de los datos. También se norma la distribución de los datos por parte del usuario.

- **Menciones y créditos:** se reconoce la propiedad intelectual; para ello se explicita la fuente de donde se obtuvieron los datos. Cualquier intercambio de datos entre investigadores no implica el otorgamiento de derechos para publicar dichos datos.

Conclusiones

Las bases de datos relacionales constituyen una herramienta clave para organizaciones e investigadores que generan grandes cantidades de información científica. Estas BD representan colecciones de datos integrados por registros, archivos y otros objetos que son almacenados y estructurados de modo que se interrelacionen y sean de fácil acceso. Una base de datos relacional facilita el ingreso, consulta, manipulación y distribución efectiva de datos a múltiples usuarios. 

Bibliografía

- Codd, E. 1970. A relational model of data for large shared databanks. dbf viewer plus 1.5. Reference manual. <http://www.programas-gratis.net/b/manual-dbf-viewer-plus-1.5>
- MySQL 5.6 Reference Manual, Including MySQL Cluster NDB 7.3 Reference Guide. [http://dev.mysql.com/doc/phpMyAdmin/Reference Manual. http://www.phpmyadmin.net/home_page/index.php](http://dev.mysql.com/doc/phpMyAdmin/ReferenceManual.html)
- PostgreSQL 8.3.3. Reference Manual. <http://www.postgresql.org/docs/8.3/interactive/>
- Quiroz, J. 2003. El modelo relacional de bases de datos. Boletín de Política Informática 6: 53-61.
- Ramos, M.J.; Ramos, A.; Montero, F. 2006. Sistemas gestores de bases de datos. España. McGrawHill. 457 p.
- Sánchez, J. 2004. Principios sobre bases de datos relacionales. Creative Commons. 32 p. <http://www.jorgesanchez.net/bd/bdrelacional.pdf>