Data Management and Archiving

A Lecture Presented to the Participants of the Seminar on
Enhancing Legislation Skills of the Regional Assembly Support Staff
Autonomous Region in Muslim Mindanao
Marco Polo Hotel, Davao City
July 22 - 24, 2009

by

DR. BRICCIO M. MERCED, JR.
University Librarian, UP Mindanao

Order of Presentation

- Definition
- Data Management Topics
- Data Administration
- Database and its Definition
- Types of Databases
- Reasons for Using a DBMS
- Sample Database Concepts
- Selecting an Appropriate DBMSystem

Definition

- is a broad field of study, but essentially is the process of managing data as a resource that is valuable to an organization or business.
- one of the largest organizations that deal with data management, DAMA (Data Management Association), states that data management is the process of developing data architectures, practices and procedures dealing with data and then executing these aspects on a regular basis.

Data Management Topics

- Data Modeling
- Data Warehousing
- Data Movement
- Database Administration
- Data Mining

Data Modeling

- is first creating a structure for the data that you collect and use and then organizing this data in a way that is easily accessible and efficient to store and pull the data for reports and analysis.
- in order to create a structure for data, it must be named appropriately and show a relationship with other data. It also must fit appropriately in a class.
- for instance, if you have a database of media, you might have a hierarchal structure of objects that include photos, videos, and audio files. Within each category, you can classify objects accordingly.

Data Warehousing

- storing data effectively so that it can be accessed and used efficiently.
- different organizations collect different types of data, but many organizations use their data the same way, in order to create reports and analyze their data to make quality business decisions.
- data warehousing is usually an organizational wide repository of data, however for very large corporations in can encompass just one office or one department.

Data Movement

- is the ability to move data from one place to another.
- for instance, data needs to be moved from where it is collected to a database and then to an end user, but this process takes quite a bit of logistic insight
- not only do all hardware, applications and data collected need to be compatible with one another, they must also be able to be classified, stored and accessed with ease within an organization
- moving data can be very expensive and can require lots of resources to make sure that data is moved efficiently, that data is secure in transit and that once it reaches the end user it can be used effectively either to be printed out as a report, saved on a computer or sent as an email attachment.

Database Administration

- is extremely important in managing data. Every organization or enterprise needs database administrators that are responsible for the database environment
- database administrators are usually given the authority to do the following tasks that include recoverability, integrity, security, availability, performance and development & testing support.

Database Recoverability

- is usually defined as a way to store data as a back up and then test the back ups to make sure that they are valid.
- the task of integrity means that data that is pulled for certain records or files are in fact valid and have high data integrity
- data integrity is extremely important especially when creating reports or when data is used for analysis. If you have data that is deemed invalid, your results will be worthless.

Database Security

- is an essential task for database administrators. For instance, database administrators are usually in charge of giving clearance and access to certain databases or trees in an organization.
- Another important task is availability. Availability is defined as making sure a database is up and running. The more up time, usually the higher level of productivity.
- Performance is related to availability, it is considered getting the most out of the hardware, applications and data as possible. Performance is usually in relation to an organizations budget, physical equipment and resources.

Involvement of a Database Administrator

- finally, a database administrator is usually involved in database development and testing support
- Database administrators are always trying to push the envelope, trying to get more use out of the data and add better performing and more powerful applications, hardware and resources to the database structure
- A database that is administered correctly is not only a sign of competent database administrator, but it also means that all end users have a huge resource in the data that is available. This makes it easy to create reports, conduct analysis and make high quality decisions based on data that is collected and used within the organization.

Data Mining

- is a process in which large amounts of data are sifted through to show trends, relationships, and patterns
- data mining is a crucial component to data management because it exposes interesting information about the data being collected
- it is important to note that data is primarily collected so it can be used to find these patterns, relationships and trends that can help a business grow or create profit
- While there are many topics within data management, they all work together from the beginning where data is collected to the end of the process where it is sifted through; analyzed and formatted where specialists can then make quality decisions based upon it.

Database

- one of the technology terms that most people have become accustomed to hearing either at work or while surfing the internet is the database
- the database used to be an extremely technical term, however with the rise of computer systems and information technology throughout our culture, the database has become a household term

Database Defined

- a database is a structured collection of records or data that is stored in a computer system
- in order for a database to be truly functional, it must not only store large amounts of records well, but be accessed easily
- in addition, new information and changes should also be fairly easy to input
- in order to have a highly efficient database system, you need to incorporate a program that manages the queries and information stored on the system
- this is usually referred to as DBMS or a Database Management System
- besides these features, all databases that are created should be built with high data integrity and the ability to recover data if hardware fails

Types of Databases

- there are several common types of databases; each type of database has its own data model (how the data is structured)
- Flat
- Hierarchical
- Network
- Relational

Flat Model Database

- In a flat model database, there is a two dimensional (flat structure) array of data
- for instance, there is one column of information and within this column it is assumed that each data item will be related to the other
- for instance, a flat model database includes only zip codes. Within the database, there will only be one column and each new row within that one column will be a new zip code.

Hierarchical Model Database

- the hierarchical model database resembles a tree like structure, such as how Microsoft Windows organizes folders and files
- in a hierarchical model database, each upward link is nested in order to keep data organized in a particular order on a same level list
- for instance, a hierarchal database of sales, may list each days sales as a separate file. Within this nested file are all of the sales (same types of data) for the day.

The Network Model Database

- in a network model, the defining feature is that a record is stored with a link to other records - in effect networked
- these networks (or sometimes referred to as pointers) can be a variety of different types of information such as node numbers or even a disk address

The Relational Model Database

- The relational model is the most popular type of database and an extremely powerful tool, not only to store information, but to access it as well
- relational databases are organized as tables. The beauty of a table is that the information can be accessed or added without reorganizing the tables. A table can have many records and each record can have many fields.

Tables in a Relational Model

- Tables are sometimes called a relation. For instance, a company can have a database called customer orders, within this database will be several different tables or relations all relating to customer orders
- tables can include customer information (name, address, contact, info, customer number, etc) and other tables (relations) such as orders that the customer previously bought (this can include item number, item description, payment amount, payment method, etc)
- it should be noted that every record (group of fields) in a relational database has its own primary key. A primary key is a unique field that makes it easy to identify a record.

Relational Database Programs

- Relational databases use a program interface called SQL or Standard Query Language
- SQL is currently used on practically all relational databases
- Relational databases are extremely easy to customize to fit almost any kind of data storage. You can easily create relations for items that you sell, employees that work for your company, etc.

Accessing Information Using a Database

- for many database users the most important feature of a database is quick and simple retrieval of information
- in a relational database, it is extremely easy to pull up information regarding an employee, but relational databases also add the power of running queries
- queries are requests to pull specific types of information and either show them in their natural state or create a report using the data. For instance, if you had a database of employees and it included tables such as salary and job description, you can easily run a query of which jobs pay over a certain amount
- no matter what kind of information you store on your database, queries can be created using SQL to help answer important questions.

Storing a Database

- databases can be very small (less than 1 MB) or extremely large and complicated (terabytes as in many government databases)
- all databases are usually stored and located on hard disk or other types of storage devices and are accessed via computer
- large databases may require separate servers and locations, however many small databases can fit easily as files located on your computer's hard drive.

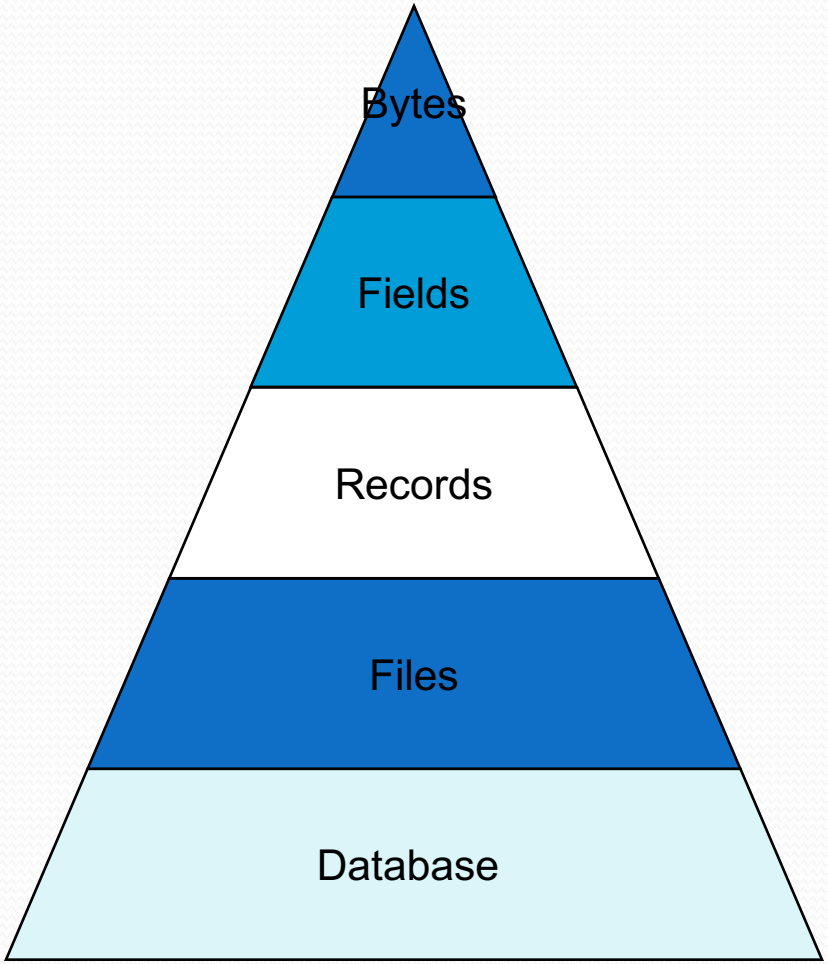
Securing a Database

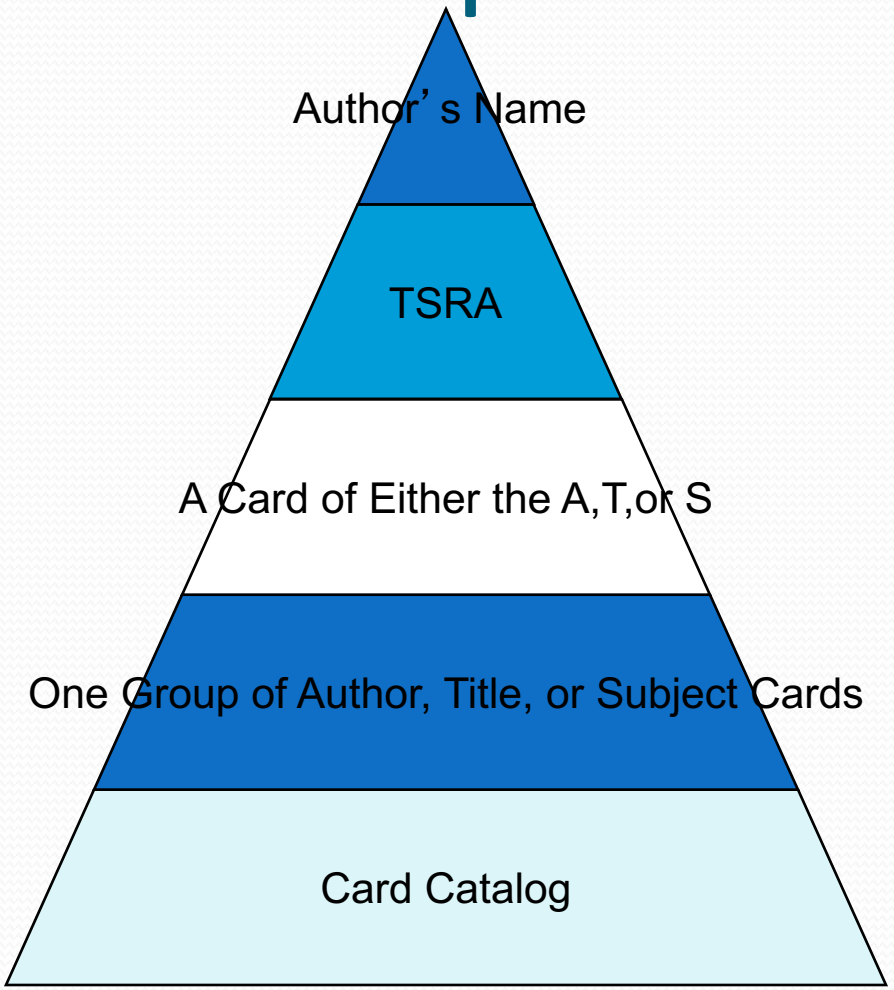
- many databases store confidential and important information that should not be easily accessed by just anyone
- many databases require passwords and other security features in order to access the information
- while some databases can be accessed via the internet through a network, other databases are closed systems and can only be accessed on site.

Reasons for Using a DBMS

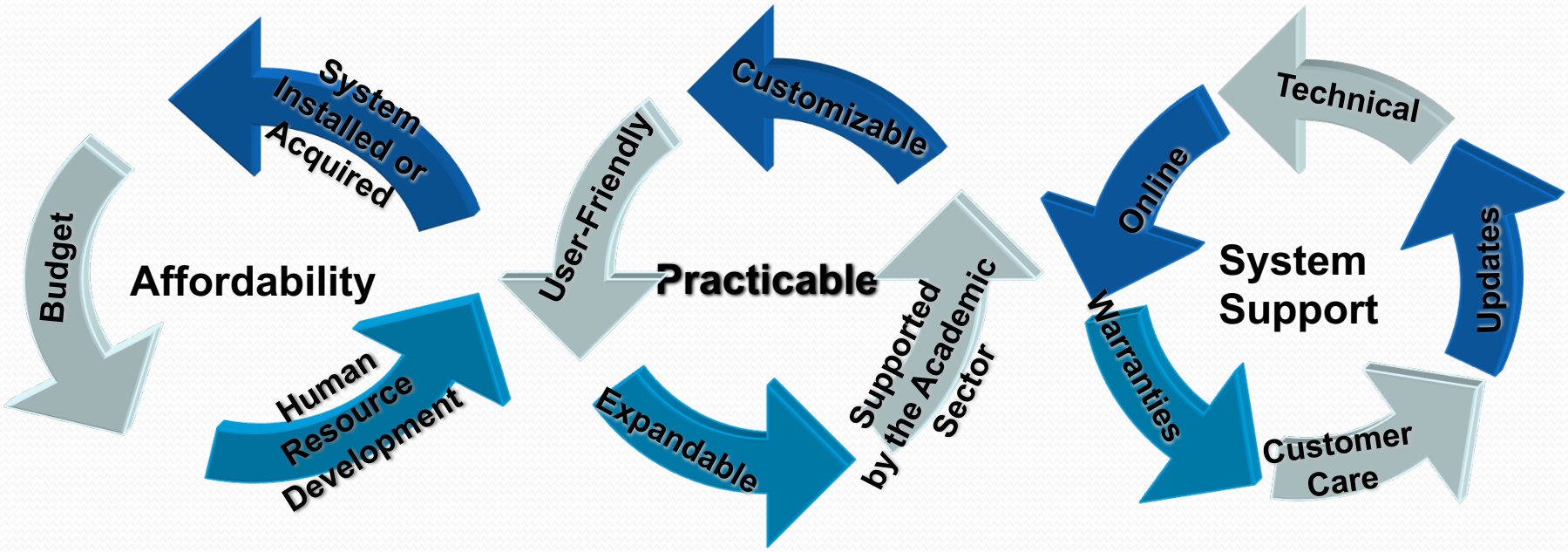
- Storing and retrieving data
- Managing metadata
- Limiting and controlling redundant data in multiple systems
- Supporting simultaneous data sharing
- Providing transaction atomicity
- Providing authorization and security services
- Enforcing business rules
- Increasing programmer productivity

Sample Database Concepts





Selecting an Appropriate System



Summarize the Key Benefits Provided by the Product

- Features of the software
- Services provided
- Online products and services
- Warranties
- Local system support (technical, etc.)
- Online support
- Pricing system (package, per module, add-ons, etc.)

References and Resources Used

- Data Management accessed at http://www.tech-faq.com/data-management.html on December 13, 2010