Databases, Data Mining & Knowledge Discovery

Charlotte Seckman, PhD, RN-BC
Assistant Professor, Course Director
University of Maryland School of Nursing
Nursing Informatics Program

Objectives

- Define key terms.
- Discuss various database models and associated terms such as data warehouse, data dashboards, data mining and knowledge discovery.

Definitions

- A Database is a shared collection of logically related data, designed to meet the information needs of multiple users in an organization
- Database Management System (DBMS): software that helps organize data in a way that allows fast and easy access to the data

Databases

Advantages

- Store large amounts of data long term
- Data easily available
- Fast response
- Simultaneous access
- Efficient exchange
- Used for research

Disadvantages

- Can be expensive to set up
- Depending on database model direct access to data may be limited
- Data excess

Database Design

From smallest to largest unit:

- Character
- Field
- Record
- File
- Database

Database Models

Types of DB Models:

- Flat
- Relational

Database Models: Flat

- Binary file; contains one record per line
- Fields separated by delimiters, e.g. commas, or may have a fixed length
- No relationship between the records
- Example: Creating a name and address list on paper, in a word processing program or a spreadsheet. This would be called a "flat file"

Example: Flat File

Last Name	First Name	Age	Gender	Race	Street Address	City	State	Zip
Bunny	Bugs	68	Male	Rabbit	555 Looney Tunes Lane	Tunetown	PA	11122
Duck	Daffy	72	Male	Duck	444 Looney Tunes Drive	Tunetown	PA	11122
Pig	Porky	69	Male	Pig	333 Looney Tunes Way	Tunetown	PA	11122
Bird	Tweety	58	Female	Bird	222 Looney Tunes Ave	Tunetown	PA	11122
Cat	Sylvester	60	Male	Cat	111 Looney Tunes Terr.	Tunetown	PA	11122
LePew	Pepe	69	Male	Skunk	110 Looney Tunes Street	Tunetown	PA	11122

Database Models: Relational

- Data is organized into two dimensional table
- Can retrieve data across all files directly without going through pathways
- Can extract key information

Database Models: Relational

- **Three Types of Rational DB**
  - Proprietary – licensed by vendors
    - All Electronic Health Record Systems
  - Open Source – freely available for use
    - MySQL, PostGIS
  - Embedded – packaged as part of other software or hardware
    - Mobile applications to store phone numbers

Relational Model Format

- Table = File or relation
- Box = Data item
- Column = Field
- Rows = Record
- Primary key = Common Field used to link Files / Tables

Relational Terms

Column=Field (Category/Name)

Results Table			
Results Number	Order Number	Hgb (gm)	Hct (%)
2082	34687	11	28
2276	58997	13	41
3488	40899	12	37
9089	21220	11	31

Relational Terms

Row=record (results)

Results Table			
Results Number	Order Number	Hgb (gm)	Hct (%)
2082	34687	11	28
2276	58997	13	41
3488	40899	12	37
9089	21220	11	31

Tables & Relationships

Patient Table		
MRN	**Last Name**	**Zip Code**
222-22-3333	Wilson	21201
111-33-5555	Smith	21234
444-55-7777	Jones	34543

"relationship"

Patient has an order

Order Table			
Order Number	**MRN**	**Order Type**	**Order Date**
34678	222-22-333	HCT/HGB	7/4/2000
23456	111-33-5555	HCT/HGB	7/4/2000
12123	444-55-7777	HCT/HGB	7/4/2000

Order has a result

"relationship"

Results Table			
Result #	**Order #**	**HGB (gm)**	**HCT%**
4567	34678	12	43
7865	23456	13	45
8942	12123	10	34

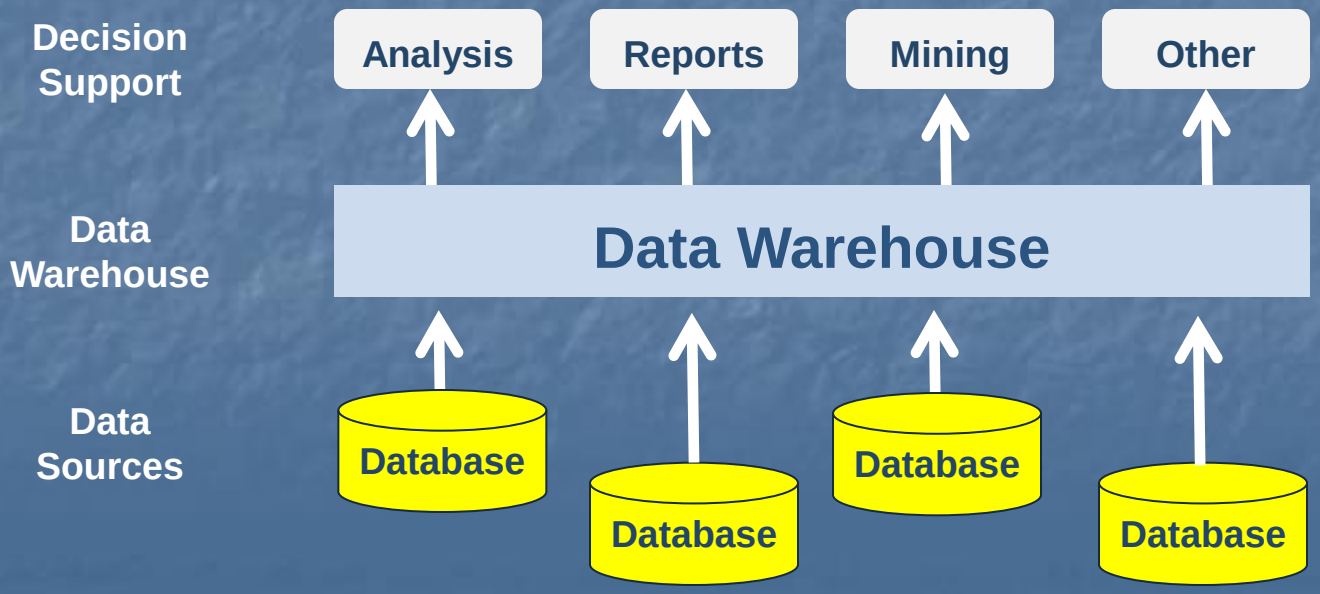
Working with Databases

- **Query** – operation to retrieve and update data from a database table i.e. ask a question
- **Reports** – report generators can filter and display selected fields from tables
- **Forms** – visual interface for user to input new data

Data Warehouse, Data Dashboard, Data Mining & Knowledge Discovery

Data Warehouse

- **Definition:**
  - Virtual storage areas for large databases
  - Contain decision support tools for analysis, reports, mining, and other processes

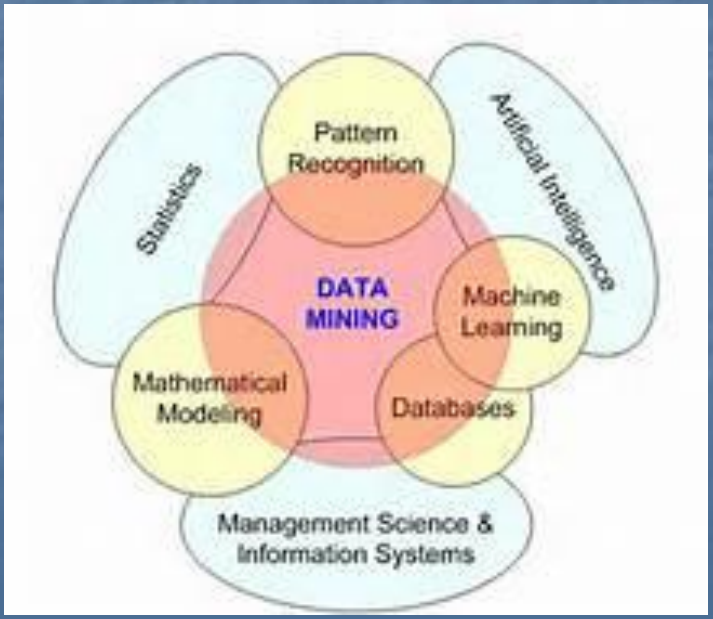


Data Dashboard

- An interface that allows users to view information from a variety of data sources simultaneously

Data Mining

- **Definition:**
  - "Technique that looks for hidden patterns and relationships in large groups of data using software" (Hebda & Czar, 2013).



Knowledge Discovery

- **Knowledge Discovery in Databases (KDD):**
  - "Extraction of implicit, unknown, and potentially useful information from data" (Hebda & Czar, 2013).
  - KDD refers to the higher level processes that include extraction, interpretation and application of data and is interrelated (and often used interchangeably) with the term data mining.

Knowledge Discovery in Databases