PDHonline Course L155G (5 PDH)

Data Models and Data processing in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

Lecture 1 Content

- ## Geographic Information Systems (GIS)

Data Models,
Data Structures and Data Management

This lecture is the beginning of GIS topics identified in the course description that is related to data models, data structures and data analysis. Having data is one aspect in the implementation of a GIS, but using it in the context of the phenomena being mapped is another aspect. This involves the phenomena being modeled, and data is structured in a manner in which the GIS software can be processed and analyzed.

Slide 2

□ **Data Base Approach**
  – **its underlying concept**
  – **advantages of the data base approach**
  – **disadvantages of the data base approach**

□ **Data Models**
  – **the understanding**
  – **records, fields and keys**

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

2

This is the content of this lecture. We will look at two topics in more detail as identified on the slide. The first topic identifies the understanding of the database approach. The second topic explains data models.

Slide 3

❑ **Database Approach**

– **a database is a collection of information about a phenomena or concept by using tables and their relationships to each other**

➢ **example people names connected to a specific address. The names and the addresses are the phenomena stored in databases**

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

3

Databases are very important in any GIS implementation. They form the basis upon which GIS processing are applied in order to get the answers to realistic questions. This means that the databases need to represent a close approximation of the phenomena being investigated. To do this, the appropriate data sets need to be collected.

Databases are varied and depend upon their intended GIS application. Databases contain data sets that are used to model the phenomena being investigated.

Slide 4

File processing approach
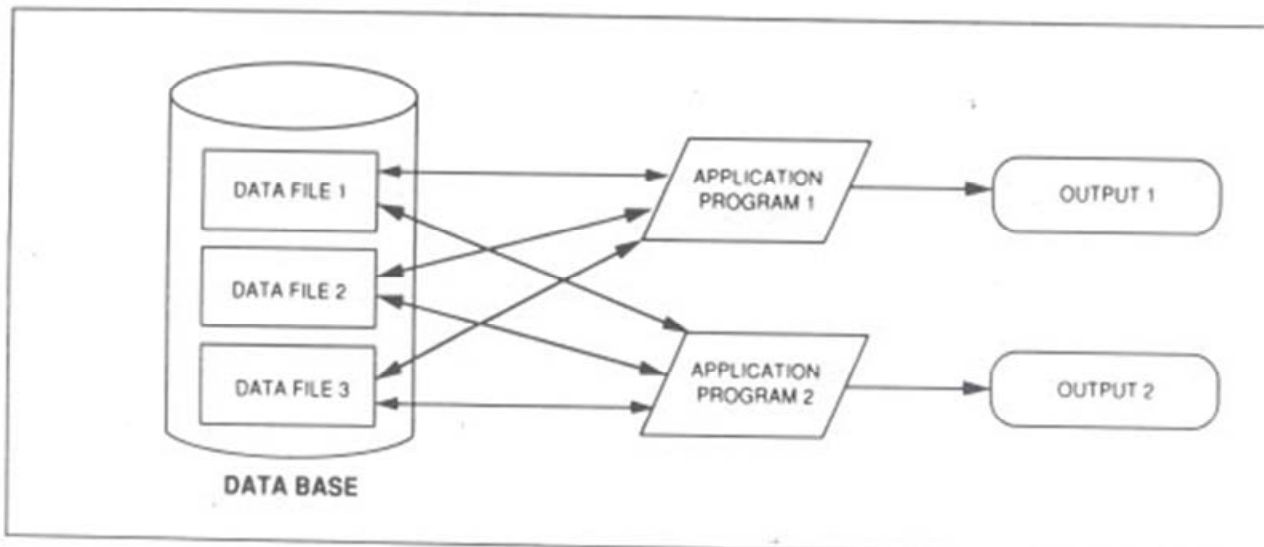
Sharing data files among applications in the file processing environment

This is one approach in which data was traditionally stored. In this approach the files are accessible by any application that is on a stand alone computer or stored on a server. Access to the various data sets is not controlled and any application can access the data files. This is approach has some obvious pros and cons. A pro is that data files are easily accessed, but at the same time the con is that, this accessibility also means that the data can be corrupted by any software application accessing it.

Slide 5

Common Data Base Approach

1. File Processing Approach

➢ **Makes use of a flat file structure to store specific characteristics that is needed to describe the given phenomena**

➢ **accessed by multiple users**

➢ **most common (but has drawbacks)**

Drawbacks:

a) **The program must know how the data in each file is stored (its characteristics, location, etc.). This can create considerable redundancy because the instructions to access a data file must be present in each application program**

b) **Problem of maintaining data integrity**

➢ **since data files can be accessed and modified by several programs and several users, then there must be some overall access control**

c) **GIS data sets are typically stored as directory or a set of related files. Since GIS data and not a single file then such approach is not appropriate**

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

5

This slide gives some self explanatory notes on the file processing approach. Even though this is common, it happens that this approach is outdated. The new method is the second method which is presented on the next slide.

2. Database Management System Environment Approach

- **A database management system (DBMS) consists of a set of programs that manipulate and maintain the data in a database.**

- **Controls the access to data in terms of the editing capability, access, making backups, and such like.**

- **Maintains data integrity which is important in any GIS implementation.**

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

6

The DBMS approach addresses the drawbacks of the file processing approach. One of the important characteristics of a DBMS is its ability to maintain data integrity and the ability to control access to the multiple data sets by various users on a network.

Slide 7

Use of a Database Management System
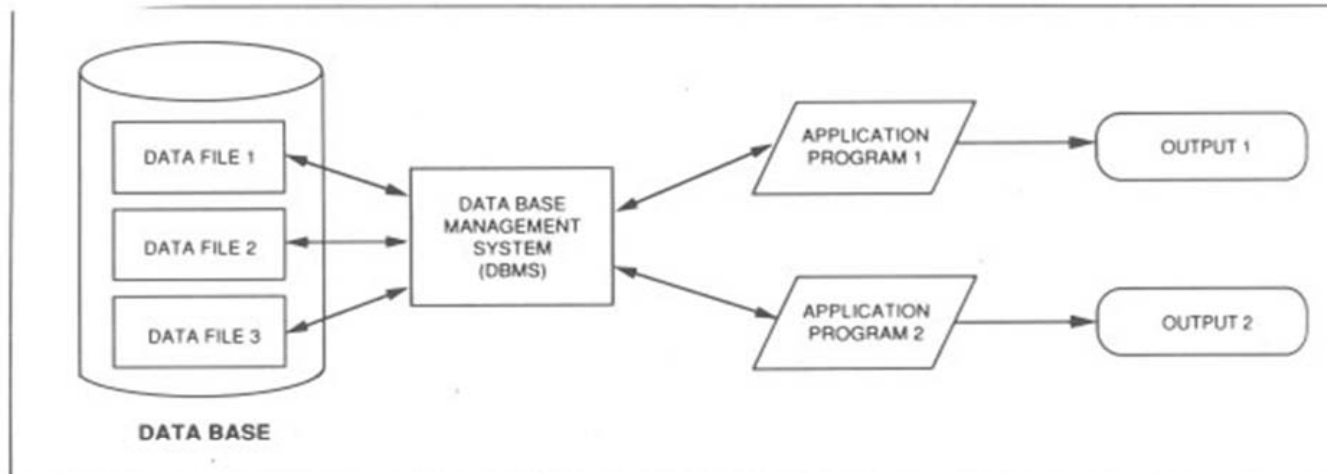
Sharing data files among applications in a database management system environment

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

7

This slide shows the system configuration and the role the DBMS regarding the data files and the application programs. The application programs interface with the DBMS to access the data files stored on a single server or multiple servers on a computer network.

Slide 8

- **DBMS were developed to manage the sharing of data in an orderly manner and to ensure that the integrity of the data base is maintained**
- **DBMS acts as a central control over all interactions between the database and the application programs, which in turn interact with the user**
- **A major benefit of a DBMS is that it provides data independence**

This slide gives further details into the DBMS approach.

Advantages of the DBMS Approach

1) *Centralized Control*

- Centralized control is made on data quality standards, security restrictions, and the integrity of the database is maintained.

2) *Data can be shared efficiently*

- Information in a database can be shared in a flexible yet controlled manner.

3) *Data Independence*

- Application programs are independent of the physical form in which the data are stored.

4) *Easier implementation of new database applications*

- Using the services provided by a DBMS, new application programs can be implemented.

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

9

This slide identifies some advantages of the DBMS approach. The information presented is self explanatory.

5) *Direct user access*

- **availability of providing a user interface so that non-programmers can perform their analyses**
- **also maintain data consistency and protect the integrity of the data base**

6) *Redundancy can be controlled*

- **data redundancy is expensive in terms of the needed additional data storage facility**
- **if redundancy is present then an effective strategy must be provided to update the multiple copies of the data**

7) *User Views*

- **a DBMS can create and maintain multiple user views**

This is a continuation of the advantages of the DBMS approach. DBMS has many advantages as they relate to GIS.

Slide 11

Disadvantages of the Data Base Management Approach

1)Cost

- **system software and any associated hardware can be expensive**
- **also an additional cost for maintenance**

2)Added Complexity

- **a DBMS is more complex than a file processing system**
- **is prone to failure and sometimes necessary to have back-ups and recovery systems available**

3)Centralized risk

- **centralizing the data and reducing data redundancy, there is a greater theoretical risk of loss or corruption of data while running an application program**

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

11

This slide identifies some disadvantages of the DBMS approach. Each point is clearly identified and explained.

Slide 12

❑Data Models

- the Real World is infinitely complex, the closer one looks, the more details is seen
- it would take an infinitely large database to capture the real world. Therefore, the data must be reduced to a finite and manageable quantity
- the rule to convert real world into discrete object is called the DATA MODEL
- Defined as :

    " A set of guidelines for the representation of the logical organization of the data in a database consisting of named logical units of data and the relationship between them."

- current GIS differ according to the way in which they organize reality through the data model
- each model tends to fit certain types of data and applications better than others

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

12

This slide addresses Data Models and the context definition of what is a data model. Put simply, a data model is a selected set of data sets which is used to represent a phenomena depicted in reality. There are various data models and such models are defined using a systematic development approach.
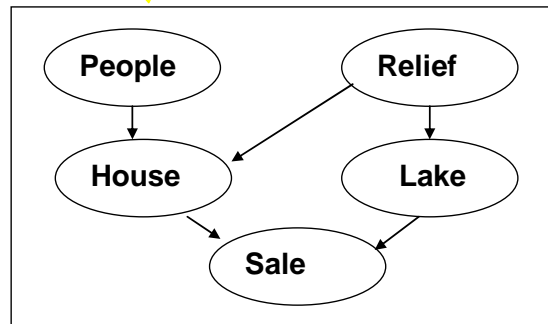
Slide 13

– **the data model chosen for a particular application is also influenced by other external factors such as:**

1. **Software available : An appropriate model must be sought which will be influenced by the software available in house, or on the market. If not, then customizing will be sought.**

2. **Training of the key individual : that is how easy can users be trained and the steepness of the learning curve**

3. **Historical precedent : involves what is familiar with the general user community**

L155 - GIS Data Models and Data Processing
Lecture 1
Dr. Steve Ramroop

13

The choice of data models are also influenced by other external factors such as the available software, training, and the historical president. Even though these are not supposed to influence the data model, they do influence the model indirectly.

Levels of modeling

Reality

Conceptual Modeling

People

Relief

House

Lake

Sale

Answer the question "What"

Logical
– **Define tables on paper**
– **Relationships**

Physical
– **Define tables on using software**

Logical / Physical Modeling

Answer the question "How"

This slide show the different levels of modeling which are defined by a conceptual model, a logical model, and a physical model.

The conceptual model identifies realistic entities and their relationships between the entities. These entities answer the question "What" is being modeled from the real world.

The Logical and Physical model answers the question "How" are the entities modeled.

The logical model defines the tables and the relationship between the tables. The relationships can be "one to one", "one to many", "many to one", or a "many to many"

The Physical model defines the column definitions for each table. That is if it is numeric, character, how many decimal places, etc. It is the Physical model that is used in defining tables representing the data model. This is done when the DBMS operator is building the database on the computer.

Note that the conceptual, and logical models are defined using diagrams, preferably using paper and pencil.

Other database terminologies:

- *Record*
  - a small group of related data items stored together
  - thought of as a row in a table
- *Entity*
  - is an object, event or concept being displayed
- *Fields*
  - sub-division of a record each of which having an item of data
- *Keys*
  - unique identifier used to reference data from the data file

This slide shows some other database terminologies. A record is a row, an entity is an event or object, a field is a column, a key is a unique identifier.

– **there are five classic data models that are used to organized electronic databases:**

    a) **Hierarchical**

    b) **Network**

    c) **Relational**

    d) **Object Oriented**

    e) **Temporal**

Details of each data models are discussed in the next lecture.

This slide identifies five classic data models which is used to the next lecture.

Slide 17

… The End …