

GIS

LECTURE 5 DATA MANAGEMENT

DATA MANAGEMENT- INTRO TO DATABASES

- What is a database?
- It is an organised collection of data.
- The data is organised to model relevant aspects of reality
- (e.g. availability of rooms in a hotel,
- finding available flights,
- finding a book to borrow in a library,
- making a tax payment entries in a computer)
- Types of Databases

DATA MANAGEMENT- INTRO TO DATABASES

- Flat file databases

ID	Name	Team
1	Idjo	A
2	Kofi	B
3	Kwame	C
4	Akos	A

- Due to the limitations of flat file databases several models have been designed over the years. These include:
- Hierarchical database
- Network database
- Relational database
- Object database
- Object-relational database (Hybrid of the relational and object) and others.

DATA MANAGEMENT- INTRO TO DATABASES

- As opposed to flat file databases,
- Relational databases allow you to define certain record fields, as keys or indexes,
- to perform search queries,
- join table records and establish integrity constraints.
- NB: Search queries are faster and more accurate when based on indexed values.

DATA MANAGEMENT- INTRO TO DATABASES

- **Data types**
- Each column in a database table such as the one above is ordinarily restricted to a specific data type.
- Double
- Integer
- Date
- Text, etc...
- **Primary and Foreign Keys**
- The primary key of a relational table uniquely identifies each record in the table.

DATA MANAGEMENT- INTRO TO DATABASES

- It can either be a normal attribute that is guaranteed to be unique (such as Social Security Number in a table with no more than one record per person).
- Primary keys may consist of a single attribute or multiple attributes in combination.
- A **foreign key** is a field in a relational table that matches the primary key of another table. The foreign key can be used to cross-reference tables.

EXAMPLE – STORING CUSTOMER INFORMATION IN A TABLE- CUSTOMER TABLE

- To uniquely identify each customer, each customer must be linked to a unique piece of information.
- This piece of information is called the **primary key**.
- For our customer contact system we will assign a unique customer ID to each customer.
- In our database design, the customer_id field will function as the primary key of the Customers table.

EXAMPLE – STORING CUSTOMER INFORMATION IN A TABLE- CUSTOMER TABLE

customer_id	first_name	last_name	email_address	telephone
0	John	Smith	john.smith@scor	123456789
1	John	Connor	j.connor@scor	987498473

orderid	customer_id	product	order_date
1	0	Samsung fridge	25-10-1989

- **Database Integrity**
- Types of integrity constraints
- Data integrity is normally enforced in a **database system** by a series of integrity constraints or rules.

DATABASE INTEGRITY

- **Entity integrity**
- Entity integrity is an integrity rule which states that every table must have a primary key- unique and not null
- **Referential integrity**
- The referential integrity rule states that any foreign key value can only be in one of two states.
- foreign key value refers to a primary key or null
- **Domain integrity**
- A domain is a set of values of the same type. Domains are therefore pools of values from which actual values appearing in the columns of a table are drawn.

DATABASES- SQL

- **What is SQL?**
- **SQL** often referred to as **Structured Query Language**, is a database computer language designed for managing **data** in relational database management systems (RDBMS), and originally based upon relational algebra and calculus.
- **What is a database schema?**
- A **database schema** of a database system is its structure described in a formal language supported by the database management system (DBMS) and refers to the organization of data to create a **blueprint of how a database will be constructed** (divided into database tables)

DATA MANAGEMENT

- **Comprises:**
 - Data Maintenance
 - Data Modelling

DATA MANAGEMENT- MAINTENANCE

- **Data Maintenance**
 - **Data maintenance** is the adding, deleting, changing and updating of **binary** and high level **files**, and the real world **data** associated with those files.
 - Data can be maintained manually and/or through an automated **program** but at origination and translation/delivery point must be translated into a binary representation for storage
 - Data is usually edited at a slightly higher level (text, images, or scientific or financial information)



DBMS

○ Data Modelling

- **Data modelling** is the process of defining real world phenomena or geographic features of interest in terms of their characteristics and their relationships with one another
- **Database administration (DBA)** is the design, implementation, maintenance and repair of an organization's database.
- **Database Management System or DBMS** is a set of computer programs that controls the creation, maintenance, and the use of a database.

DBMS

- A DBMS is a system software package that helps the use of integrated collection of data records and files known as databases.
- It allows different user application programs to easily access the same database.
- user can ask simple questions in a query language (SQL- Structured Query Language)
- many DBMS packages provide Fourth-generation programming language (4GLs, 1970s-1990)
- DBMS helps specify the logical organization for a database and access and use the information within a database

DBMS

- It provides facilities for
 - controlling data access,
 - enforcing data integrity,
 - managing concurrency, and
 - restoring the database from backups
- A DBMS also provides the ability to logically present database information to users
- E.g MySQL, Sybase, Oracle, MS Access etc etc

DATA MODELLING- DBMS

- Data modelling - defining real world phenomena or geographic features of interest in terms of their characteristics and their relationships with one another
 - it is concerned with different phases of work carried out to implement information organization and data structure
- There are three steps in the data modelling process,
 - *conceptual data modelling* --- defining in broad and generic terms the scope and requirements of a database
 - *logical data modelling* --- specifying the user's view of the database with a clear definition of attributes and relationships
 - *physical data modelling* --- specifying internal storage structure and file organization of the database

DATA MODELLING- DBMS

- Data modelling is obviously closely related(synonymous) to the three levels of data abstraction in database design:
 - conceptual data modelling ----> data model
 - logical data modelling -----> data structure
 - physical data modelling -----> file structure

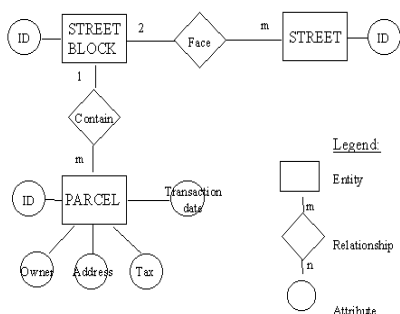
CONCEPTUAL DATA MODELLING

A method of *semantic data modelling* =human language-like vocabulary to describe information organization

It involves four aspects of work:

- identifying entities
 - an entity is defined as a person, a place, an event, a thing, etc.
- identifying attributes
- determining relationships
- Drawing an *entity-relationship diagram* (E-R diagram)

ENTITY-RELATIONAL DIAGRAM



A.K. Yeung 1998-10-10 v51-21

E-R DIAGRAM

- A street block contains parcels of land
- A street block faces many streets
- Parcels have IDs, an owner, an address and a tax to be paid

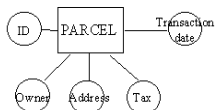
LOGICAL DATA MODELLING

- Logical data modelling is a comprehensive process by which the conceptual data model is consolidated and refined
- The proposed database is reviewed in its entirety in order to identify potential problems such as
 - irrelevant data that will not be used
 - omitted or missing data
 - inappropriate representation of entities
 - lack of integration between various parts of the database
 - unsupported applications
 - potential additional cost to revise the database

The end product of logical data modelling is a *logical schema*

- the logical schema is developed by mapping the conceptual data model (such as the E-R diagram) to a software-dependent design document

Logical schema of entity PARCEL



PARCEL (Identifier, Owner, Address, Tax, Transaction date)

Relation name
Primary key

A.K. Yeung 1998-10-10 v51-22

PHYSICAL DATA MODELLING

- Physical data modelling is the database design process by which the actual tables that will be used to store the data are defined in terms of
 - data format --- the format of the data that is specific to a database management system (DBMS)
 - storage requirements --- the volume of the database
 - physical location of data --- optimizing system performance by minimizing the need to transmit data between different storage devices or data servers
- The end product of physical data modelling is a *physical schema*
 - a physical schema is also variably known as *data dictionary*, *item definition table*, *data specific table* or *physical database definition*
 - it is both software- and hardware specific (this means the physical schemas for different systems look different from one another)

Example of a physical schema

Item Definition Table:

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	NDEC	ALT NAME
1	PARCEL-ID	10	15	F	-	-
11	AREA	10	15	F	2	-
21	LUL-CLASS	3	5	C	-	-
24	OWNER-LN	15	20	C	-	-
29	OWNER-FN	15	20	C	-	-
44	ADDRESS-1	30	35	C	-	-
74	ADDRESS-2	30	35	C	-	-
104	TRNS-DATE	8	10	E	-	-
112	ASS-VALUE	10	15	I	-	-
122	TAX-RATE	5	10	F	3	-
127	TAX	8	10	F	2	-

Explanatory notes:

PARAMETER	DESCRIPTION
Item name	Any name to 16 characters
Width	No. of space used to store item values
Output	No. of space used to display item values
Type	Data item type:
C	Character
I	Integer
B	Binary
N	Number
D	Date mm/dd/yy
F	Floating point
N DEC	No. of decimal points

A.K. Yeung 1998-10-10 u51-23

