**ITEC 630**

*Information Systems Analysis, Modeling, and Design*

***Lecture Notes***

**Designing Databases**

Created by Daniel LeRevised by Daniel Le 08/01/2016

**Learning objectives**

1. Understand database design concepts
2. Learn the relational database model
3. Learn how to use normalization to efficiently store data in a database
4. Learn how to transform E-R diagrams into relations
5. Learn how to design physical files and databases
6. Understand the role of the data base administrator
7. Understand the relationship of business intelligence to data warehouses, big data, business analytics and text analytics in helping systems and people make decisions.

**Overview**

During the system analysis phase, data flow diagrams are used to create a logical design for the information system. In this systems design phase, a physical design for data organization, storage, and retrieval will be developed for the system. This week lecture presents database concepts, discusses file-based systems and database systems, shows how to create entity-relationship diagrams, and concludes with a discussion of data warehouses and data mining. We will learn guidelines for well-structured and efficient database files and about logical and physical database design.

* ***Introduction***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=991460443&rowid=672&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=991460443&rowid=672&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

***Note #1: All links provided in this lecture can be activated with a "Ctrl + Click"; however, you can also activate these links by copy and paste the link content to the Web browser address bar, just in case.***

***Note #2: To access links associated with "http://library.books24x7.com.ezproxy.umuc.edu", you might have to log into UMUC Library and activate the link "Books24x7" first.***

## DATABASE DESIGN & RELATIONAL DATABASE MODEL

* Primary key is an attribute whose value is unique across all occurrences of a relation.
* Relational database model is data represented as a set of related tables or relations.
* Well-structured relation (or table) is a relation that contains a minimum amount of redundancy and allows users to insert, modify, and delete the rows without errors or inconsistencies.
* ***Database Design***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=412627118&rowid=676&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=412627118&rowid=676&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

* ***Relational Database Model***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=425820518&rowid=688&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=425820518&rowid=688&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

**NORMALIZATION**

* Normalization is the process of converting complex data structures into simple, stable data structures.
* Functional dependency is a particular relationship between two attributes.
* Second normal form (2NF) A relation for which every nonprimary key attribute is functionally dependent on the whole primary key.
* Third normal form (3NF) A relation that is in second normal form and that has no functional (transitive) dependencies between two (or more) nonprimary key attributes.
* ***Normalization***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=810122485&rowid=695&noteMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=810122485&rowid=695&noteMenuToggle=0&leftMenuState=1)

***Three steps of data normalization***

A. Remove all repeating groups and identify the primary key.

* First Normal Form (lNF).

B. Ensure that all non-key attributes are fully dependent on the primary key.   
 Remove partial key dependencies.

* Second Normal Form (2NF).

C. Remove any transitive dependencies, attributes that are dependent on  
other non-key attributes. Remove non-key dependencies.

* Third Normal Form (3NF).
* ***Develop ERD***

[**https://learn.saylor.org/mod/resource/view.php?id=12352**](https://learn.saylor.org/mod/resource/view.php?id=12352)

**Click “The New Software Engineering.pdf” link to view the file.**

**Read the “Table 9-1 Normalization Rules” section in Chapter 9 (pages 344-346).**

**TRANSFORMING E-R DIAGRAMS INTO RELATIONS**

* Foreign key is an attribute that appears as a nonprimary key attribute in one relation and as a primary key attribute (or part of a primary key) in another relation.
* Referential integrity An integrity constraint specifying that the value (or existence) of an attribute in one relation depends on the value (or existence) of the same attribute in another relation.
* Recursive foreign key is a foreign key in a relation that references the primary key values of that same relation.
* ***Transforming E-R diagrams into Relations***

## <http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=162892429&rowid=705&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1>

**MERGING RELATIONS**

* Synonyms are two different names that are used for the same attribute.
* Homonym is a single attribute name that is used for two or more different attributes.
* ***Merging Relations***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=172215639&rowid=725&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=172215639&rowid=725&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

**PHYSICAL FILE AND DATABASE DESIGN**

* A field is the smallest unit of named application data recognized by system software.
* A data type is a coding scheme recognized by system software for representing organizational data.
* Calculated (or computed or derived) field is a field that can be derived from other database fields.
* A default value is a value a field will assume unless an explicit value is entered for that field.
* A null value is a special field value, distinct from a zero, blank, or any other value, that indicates that the value for the field is missing or otherwise unknown.
* A physical table is a named set of rows and columns that specifies the fields in each row of the table.
* Denormalization is the process of splitting or combining normalized relations into physical tables based on affinity of use of rows and fields.
* A physical file is a named set of table rows stored in a contiguous section of secondary memory.
* File organization is a technique for physically arranging the records of a file.
* A pointer is a field of data that can be used to locate a related field or row of data.
* In a sequential file organization the rows in the file are stored in sequence according to a primary key value.
* In an indexed file organization the rows are stored either sequentially or nonsequentially and an index is created that allows software to locate individual rows.
* An index is a table used to determine the location of rows in a file that satisfy some condition.
* Secondary key is one or a combination of fields for which more than one row may have the same combination of values.
* In a hashed file organization the address for each row is determined using an algorithm.
* ***Physical File and Database Design***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=122467293&rowid=731&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=122467293&rowid=731&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

* ***Designing Fields***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=671223004&rowid=732&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=671223004&rowid=732&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

* ***Designing Physical Tables***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=347482134&rowid=740&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=347482134&rowid=740&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

**THE ROLE OF THE DATA BASE ADMINISTRATOR**

* Managing the data base requires a data base administrator (DBA) whose key functions are to manage data activities, the data base structure, and the DBMS.
* In addition to a managerial background, the DBA needs technical knowledge to deal with data base designers. Important for the success of this important job is the support of the senior MIS staff and upper management for the overall data base function.
* ***The Role of the Data Base Administrator***

[**http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=334947744&rowid=768&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1**](http://library.books24x7.com.ezproxy.umuc.edu/assetviewer.aspx?bookid=30713&chunkid=334947744&rowid=768&noteMenuToggle=0&hitSectionMenuToggle=0&leftMenuState=1)

**DATA WAREHOUSES, BIG DATA, BUSINESS ANALYTICS  
AND TEXT ANALYTICS**

Data warehouses are used organize information for handling queries quickly and effectively. There are many differences between data warehouse and database. When data sets are too big and/or too complex then big data is used to handle it. In order to make use of available data, techniques such as business analytics and text analytics are used to extract and/or convert data into useful information.

* ***Data-warehouses***

[**http://www.w3computing.com/systemsanalysis/data-warehouses/**](http://www.w3computing.com/systemsanalysis/data-warehouses/)

**References**

1. Conger, Sue (2008). The New Software Engineering. A Creative Commons Attribution 3.0 License.
2. J.B. Dixit and Raj Kumar (2007). Structured System Analysis and Design, Laxmi Publications. <http://library.books24x7.com.ezproxy.umuc.edu/toc.aspx?bookid=30713>
3. <http://www.w3computing.com/systemsanalysis/data-warehouses/>
4. <https://www.studytonight.com/dbms/database-normalization.php>

**Disclaimer:** Some articles or sites I select or refer you to may include materials or opinions associated with particular political or other ideological positions. My pointing to these sites in no way suggests that I am encouraging a particular ideology or position on this or any other related topic; these are simply some of the more interesting and informative sites I have found to cover the topic(s) at hand. As always, you must consider the source when reading any materials that may reflect a particular ideological point of view on an issue.

<https://www.studytonight.com/dbms/database-normalization.php>