




Modern Systems Analysis and Design

Fifth Edition

**Jeffrey A. Hoffer
Joey F. George
Joseph S. Valacich**

Chapter 10

Designing Databases



Learning Objectives

- ✓ Concisely define each of the following key database design terms: relation, primary key, normalization, functional dependency, foreign key, referential integrity, field, data type, null value, denormalization, file organization, index, and secondary key.
- ✓ Explain the role of designing databases in the analysis and design of an information system.
- ✓ Transform an entity-relationship (E-R) diagram into an equivalent set of well-structured (normalized) relations.

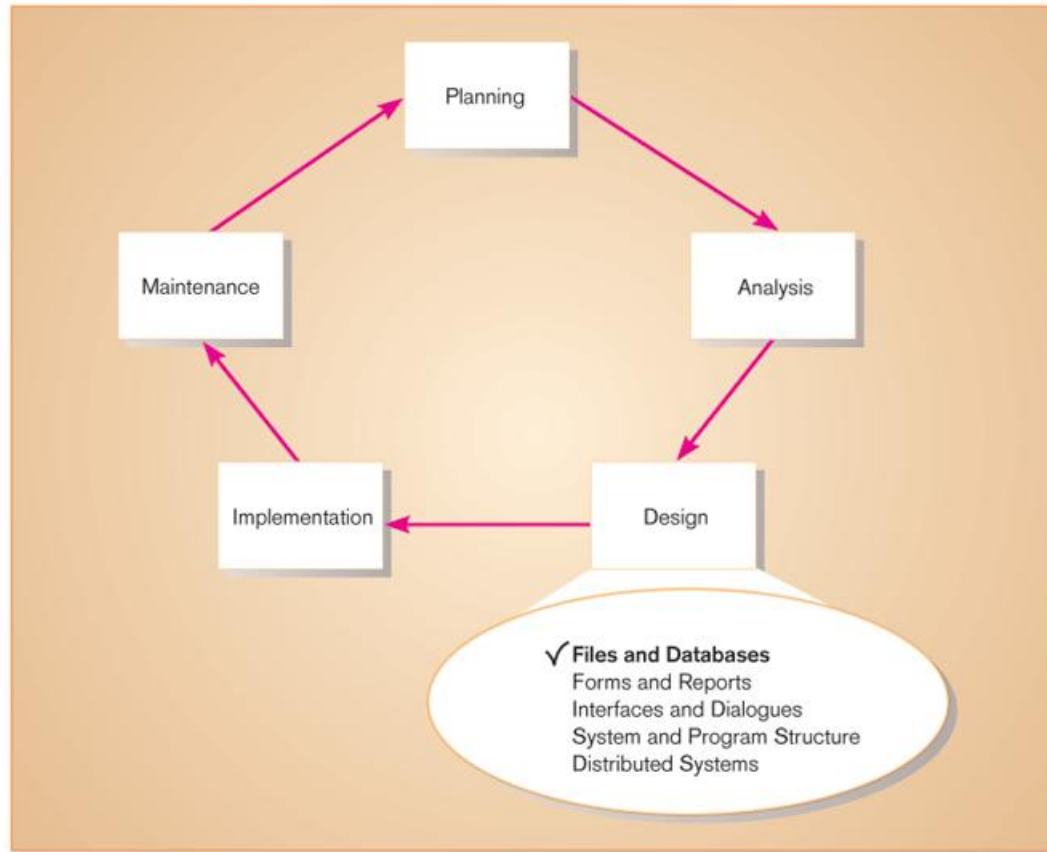


Learning Objectives (Cont.)

- ✓ Merge normalized relations from separate user views into a consolidated set of well-structured relations.
- ✓ Choose storage formats for fields in database tables.
- ✓ Translate well-structured relations into efficient database tables.
- ✓ Explain when to use different types of file organizations to store computer files.
- ✓ Describe the purpose of indexes and the important considerations in selecting attributes to be indexed.

Introduction

Figure 10-1 Systems development life cycle with design phase highlighted



Database Design

- File and database design occurs in two steps.
- Develop a logical database model, which describes data using notation that corresponds to a data organization used by a database management system.
 - Relational database model.
- Prescribe the technical specifications for computer files and databases in which to store the data.
 - Physical database design provides specifications.
- Logical and physical database design in parallel with other system design steps.

The Process of Database Design (Cont.)

- Four key steps in logical database modeling and design:
 - Develop a logical data model for each known user interface for the application using normalization principles.
 - Combine normalized data requirements from all user interfaces into one consolidated logical database model (view integration).
 - Translate the conceptual E-R data model for the application into normalized data requirements.
 - Compare the consolidated logical database design with the translated E-R model and produce one final logical database model for the application.

Physical Database Design

- Key physical database design decisions include:
 - Choosing storage format for each attribute from the logical database model.
 - Grouping attributes from the logical database model into physical records.
 - Arranging related records in secondary memory (hard disks and magnetic tapes) so that records can be stored, retrieved and updated rapidly.
 - Selecting media and structures for storing data to make access more efficient.



Deliverables and Outcomes

- Logical database design
 - Must account for every data element on a system input or output.
 - Normalized relations are the primary deliverable.
- Physical database design
 - Convert relations into database tables.
 - Programmers and database analysts code the definitions of the database.
 - Written in Structured Query Language (SQL).



Relational Database Model

- **Relational database model:** data represented as a set of related tables or relations.
- **Relation:** a named, two-dimensional table of data. Each relation consists of a set of named columns and an arbitrary number of unnamed rows.

Relational Database Model (Cont.)

- Relations have several properties that distinguish them from nonrelational tables:
 - Entries in cells are simple.
 - Entries in columns are from the same set of values.
 - Each row is unique.
 - The sequence of columns can be interchanged without changing the meaning or use of the relation.
 - The rows may be interchanged or stored in any sequence.

Well-Structured Relation and Primary Keys

- **Well-Structured Relation (or table)**
 - A relation that contains a minimum amount of redundancy;
 - Allows users to insert, modify, and delete the rows without errors or inconsistencies.
- **Primary Key**
 - An attribute whose value is unique across all occurrences of a relation.
- All relations have a primary key.
- This is how rows are ensured to be unique.
- A primary key may involve a single attribute or be composed of multiple attributes.

Normalization and Rules of Normalization

- **Normalization:** the process of converting complex data structures into simple, stable data structures.
- **First Normal Form (1NF)**
 - Unique rows, no multivalued attributes.
 - All relations are in 1NF.
- **Second Normal Form (2NF)**
 - Each nonprimary key attribute is identified by the whole key (called full functional dependency).

Rules of Normalization (Cont.)

- Third Normal Form (3NF)
 - Nonprimary key attributes do not depend on each other (i.e. no transitive dependencies).
- The result of normalization is that every nonprimary key attribute depends upon the whole primary key.

Functional Dependencies and Primary Keys

■ Functional Dependency

- A particular relationship between two attributes.
- For a given relation, attribute B is functionally dependent on attribute A if, for every valid value of A, that value of A uniquely determines the value of B.
- The functional dependence of B on A is represented by $A \rightarrow B$.

Functional Dependencies and Primary Keys (Cont.)

- Functional dependency is not a mathematical dependency.
- Instances (or sample data) in a relation do not prove the existence of a functional dependency.
- Knowledge of problem domain is most reliable method for identifying functional dependency.

Second Normal Form (2NF)

- A relation is in second normal form (2NF) if any of the following conditions apply:
 - The primary key consists of only one attribute.
 - No nonprimary key attributes exist in the relation.
 - Every nonprimary key attribute is functionally dependent on the full set of primary key attributes.
- To convert a relation into 2NF, you decompose the relation into new relations using the attributes, called *determinants*, that determine other attributes.
- The determinants are the primary key of the new relation.



Third Normal Form (3NF)

- A relation is in third normal form (3NF) if it is in second normal form (2NF) and there are no functional (transitive) dependencies between two (or more) nonprimary key attributes.

Third Normal Form (3NF) (Cont.)

- **Foreign Key:** 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.

Transforming E-R Diagrams into Relations

- It is useful to transform the conceptual data model into a set of normalized relations.
- Steps
 - *Represent entities.*
 - *Represent relationships.*
 - *Normalize the relations.*
 - *Merge the relations.*



Representing Entities

- Each regular entity is transformed into a relation.
- The identifier of the entity type becomes the primary key of the corresponding relation.



Representing Entities

- The primary key must satisfy the following two conditions.
 - The value of the key must uniquely identify every row in the relation.
 - The key should be nonredundant.
- The entity type label is translated into a relation name.

Binary 1:N and 1:1 Relationships

- The procedure for representing relationships depends on both the degree of the relationship – unary, binary, ternary – and the cardinalities of the relationship.
- **Binary 1:N Relationship:** is represented by adding the primary key attribute (or attributes) of the entity on the one side of the relationship as a foreign key in the relation that is on the many side of the relationship.

Binary 1:N and 1:1 Relationships (Cont.)

- **Binary or Unary 1:1 Relationship:**
represented by any of the following choices:
 - Add the primary key of A as a foreign key of B.
 - Add the primary key of B as a foreign key of A.
 - Both of the above.

Binary and Higher-Degree M:N Relationships (Cont.)

- **Binary and Higher-Degree M:N relationships**
 - Create another relation and include primary keys of all relations as primary key of new relation.

Unary Relationships

■ **Unary 1:N Relationship**

- Is modeled as a relation.
- Primary key of that relation is the same as for the entity type.
- Foreign key is added to the relation that references the primary key values.

- **Recursive foreign key:** A foreign key in a relation that references the primary key values of that same relation.

Unary Relationships

- **Unary M:N Relationship**
- Is modeled as one relation.
- Create a separate relation the represent the M:N relationship.
- Primary key of new relation is a composite key of two attributes that both take their values from the same primary key.
- Any attribute associated with the relationship is included as a nonkey attribute in this new relation.



Merging Relations

- Purpose is to remove redundant relations.
- The last step in logical database design.
- Prior to physical file and database design.



View Integration Problems

- Must understand the meaning of the data and be prepared to resolve any problems that arise in the process.
- **Synonyms:** two different names used for the same attribute.
 - When merging, get agreement from users on a single, standard name.

View Integration Problems (Cont.)

- **Homonyms:** a single attribute name that is used for two or more different attributes.
 - Resolved by creating a new name.
- **Dependencies between nonkeys:** dependencies may be created as a result of view integration.
 - In order to resolve, the new relation must be normalized.




View Integration Problems (Cont.)

- **Class/Subclass:** relationship may be hidden in user views or relations.
 - Resolved by creating a new name.



Physical File and Database Design

- The following information is required:
 - Normalized relations, including volume estimates.
 - Definitions of each attribute.



Physical File and Database Design (Cont.)

- Descriptions of where and when data are used, entered, retrieved, deleted, and updated (including frequencies).
- Expectations or requirements for response time and data integrity.
- Descriptions of the technologies used for implementing the files and database.

Designing Fields (Cont.)

- **Field:** the smallest unit of named application data recognized by system software.
 - Attributes from relations will be represented as fields.
- **Data Type:** a coding scheme recognized by system software for representing organizational data.

Choosing Data Types

- Selecting a data type balances four objectives:
 - Minimize storage space.
 - Represent all possible values of the field.
 - Improve data integrity of the field.
 - Support all data manipulations desired on the field.

Calculated Fields

- **Calculated (or computed or derived) field:** a field that can be derived from other database fields.
- It is common for an attribute to be mathematically related to other data.
- The calculate value is either stored or computed when it is requested.

Controlling Data Integrity

- **Default Value:** a value a field will assume unless an explicit value is entered for that field.
- **Range Control:** limits range of values that can be entered into field.
 - Both numeric and alphanumeric data.
- **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.
- **Null Value:** a special field value, distinct from zero, blank, or any other value, that indicates that the value for the field is missing or otherwise unknown.



Designing Physical Tables

- Relational database is a set of related tables.
- **Physical Table:** a named set of rows and columns that specifies the fields in each row of the table.
- **Denormalization:** the process of splitting or combining normalized relations into physical tables based on affinity of use of rows and fields.
- *Denormalization* optimizes certain data processing activities at the expense of others.

Designing Physical Tables (Cont.)

- Three types of table partitioning:
- *Range partitioning*: partitions are defined by nonoverlapping ranges of values for a specified attribute.
- *Hash partitioning*: a table row is assigned to a partition by an algorithm and then maps the specified attribute value to a partition.
- *Composite partitioning*: combines range and hash partitioning by first segregating data by ranges on the designated attribute, and then within each of these partitions.

Designing Physical Tables (Cont.)

- Various forms of *denormalization*, which involves combining data from several normalized tables, can be done.
 - No hard-and-fast rules for deciding.
- Three common situations where *denormalization* may be used:
 - *Two entities with a one-to-one relationship.*
 - *A many-to-many relationship (associative entity) with nonkey attributes.*
 - *Reference data.*



File Organizations

- **File organization:** a technique for physically arranging the records of a file.
- **Physical file:** a named set of table rows stored in a contiguous section of secondary memory.

File Organizations (Cont.)

- **Sequential file organization:** a file organization in which rows in a file are stored in sequence according to a primary key value.
- **Hashed file organization:** a file organization in which the address for each row is determined using an algorithm.
- **Pointer:** a field of data that can be used to locate a related field or row of data.

Arranging Table Rows (Cont.)

- Objectives for choosing file organization
 - Fast data retrieval.
 - High throughput for processing transactions.
 - Efficient use of storage space.
 - Protection from failures or data loss.
 - Minimizing need for reorganization.
 - Accommodating growth.
 - Security from unauthorized use.
 - Protection from failures or data loss.
 - Minimizing need for reorganization.
 - Accommodating growth.
 - Security from unauthorized use.



Indexed File Organization

- **Indexed file organization:** a file organization in which rows are stored either sequentially or nonsequentially, and an index is created that allows software to locate individual rows.
- **Index:** a table used to determine the location of rows in a file that satisfy some condition.
- **Secondary keys:** one or a combination of fields for which more than one row may have the same combination of values.

Indexed File Organization (Cont.)

- Main disadvantages are:
 - Extra space required to store the indexes; and
 - Extra time necessary to access and maintain indexes.
- Main advantages are:
 - Allows for both random and sequential processing.
- Guidelines for choosing indexes:
 - Specify a unique index for the primary key of each table.
 - Specify an index for foreign keys.
 - Specify an index for nonkey fields that are referenced in qualification, sorting and grouping commands for the purpose of retrieving data.

Designing Controls for Files

- Two of the goals of physical table design are *protection from failure or data loss* and *security from unauthorized use*.
- These goals are achieved primarily by implementing controls on each file.
- Two other important types of controls *address file backup* and *security*.

Designing Controls for Files (Cont.)

- Techniques for file restoration include:
 - Periodically making a backup copy of a file.
 - Storing a copy of each change to a file in a transaction log or audit trail.
 - Storing a copy of each row before or after it is changed.
- Build data security into a file include:
 - Coding, or encrypting, the data in the file.
 - Requiring data file users to identify themselves by entering user names and passwords.
 - Prohibiting users from directly manipulating any data in the file by forcing users to work with a copy (real or virtual).

Physical Database Design for Hoosier Burger

- The following decisions need to be made:
 - Create one or more fields for each attribute and determine a data type for each field.
 - For each field, decide if it is calculated; needs to be coded or compressed; must have a default value or picture; or must have range, referential integrity, or null value controls.
 - For each relation, decide if it should be denormalized to achieve desired processing efficiencies.
 - Choose a file organization for each physical file.
 - Select suitable controls for each file and the database.



Electronic Commerce Application: Designing Databases

- Designing databases for Pine Valley Furniture's WebStore
 - Review the conceptual model (E-R diagram).
 - Examine the lists of attributes for each entity.
 - Complete the database design.
 - Share all design information with project team to be turned into a working database during implementation.

Summary

- In this chapter you learned how to:
 - ✓ Concisely define each of the following key database design terms: relation, primary key, normalization, functional dependency, foreign key, referential integrity, field, data type, null value, denormalization, file organization, index, and secondary key.
 - ✓ Explain the role of designing databases in the analysis and design of an information system.
 - ✓ Transform an entity-relationship (E-R) diagram into an equivalent set of well-structured (normalized) relations.

Summary (Cont.)

- ✓ Merge normalized relations from separate user views into a consolidated set of well-structured relations.
- ✓ Choose storage formats for fields in database tables.
- ✓ Translate well-structured relations into efficient database tables.
- ✓ Explain when to use different types of file organizations to store computer files.
- ✓ Describe the purpose of indexes and the important considerations in selecting attributes to be indexed.