Universitas Esa Unggul

Smart, Creative and Entrepreneurial

CCA220-Analisis dan Perancangan system Informasi

www.esaunggul.ac.id

Dosen Pengampu :

5165-Kundang K Juman, Ir, MMSI

Prodi Teknik Informatika dan Sistem Informasi - Fakultas Ilmu Komputer

Chapter 13
Designing Databases

Systems Analysis and Design

Kendall & Kendall

Sixth Edition

Major Topics

- Files
- Databases
- Normalization
- Key design
- Using the database
- Data warehouses
- Data mining

© 2005 Pearson Prentice Hall

Data Storage Design Objectives

The objectives in the design of data storage organization are:

- The data must be available when the user wants to use it.
- The data must have integrity.
  - It must be accurate and consistent.
- Efficient storage of data as well as efficient updating and retrieval.

© 2005 Pearson Prentice Hall

Data Storage Design Objectives

- The objectives in the design of data storage organization are (continued):
  - The information retrieval be purposeful.
  - The information obtained from the stored data must be in an integrated form to be useful for:
    - Managing.
    - Planning.
    - Controlling.
    - Decision making.

© 2005 Pearson Prentice Hall

Approaches to Data Storage

There are two approaches to the storage of data in a computer system:

- Store the data in individual files each unique to a particular application.
- Storage of data in a computer-based system involves building a database.
  - A database is a formally defined and centrally controlled store of data intended for use in many different applications.

© 2005 Pearson Prentice Hall

Files

- A file can be designed and built quite rapidly, and the concerns for data availability and security are minimized.
- Analysts can choose an appropriate file structure according to the required processing speed of the particular application system.

© 2005 Pearson Prentice Hall

Objectives of Effective Databases

The effectiveness objectives of the database include:

- Ensuring that data can be shared among users for a variety of applications.
- Maintaining data that are both accurate and consistent.
- Ensuring all data required for current and future applications will be readily available.

© 2005 Pearson Prentice Hall

Objectives of Effective Databases

- The effectiveness objectives of the database include (continued):
  - Allowing the database to evolve and the needs of the users to grow.
  - Allowing users to construct their personal view of the data without concern for the way the data are physically stored.

© 2005 Pearson Prentice Hall

Metadata

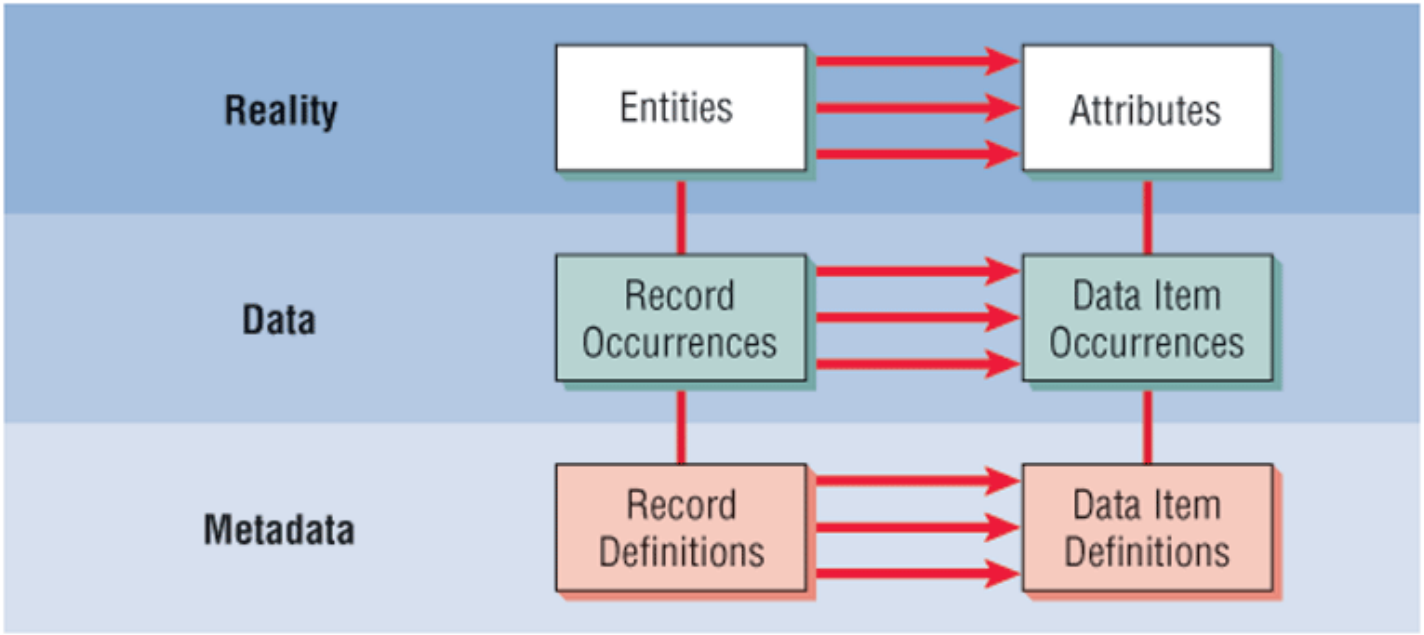
Metadata is the information that describes data in the file or database.

- Used to help users understand the form and structure of the data

© 2005 Pearson Prentice Hall

Reality, Data, and Metadata

Figure 13.1 Reality, data, and metadata.



© 2005 Pearson Prentice Hall

Entity-Relationship Concepts

- Entities are objects or events for which data is collected and stored.
- An entity subtype represents data about an entity that may not be found on every record.
- Relationships are associations between entities.

© 2005 Pearson Prentice Hall

Entities

A distinct collection of data for one person, place, thing, or event.

© 2005 Pearson Prentice Hall

Entity Subtype

- An entity subtype is a special one-to-one relationship used to represent additional attributes, which may not be present on every record of the first entity.
- This eliminates null fields on the primary database.
- For example, a company that has preferred customers, or student interns may have special field.

© 2005 Pearson Prentice Hall

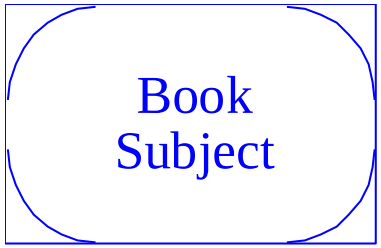
Associative Entity

- Associative Entity - links two entities
- An associative entity can only exist between two entities

© 2005 Pearson Prentice Hall

Attributive Entity

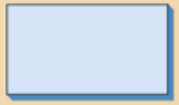
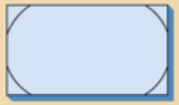
- An attributive Entity - describes attributes, especially repeating elements.



© 2005 Pearson Prentice Hall

Entity-Relationship Diagram Symbols

Figure 13.3 The entity-relationship symbols and their meanings.

Symbol	Official Explanation	What It Really Means
	Entity	A class of persons, places, or things
	Associative entity	Used to join two entities
	Attributive entity	Used for repeating groups
	To 1 relationship	Exactly one
	To many relationship	One or more
	To 0 or 1 relationship	Only zero or one
	To 0 or more relationship	Can be zero, one, or more
	To more than 1 relationship	Greater than one

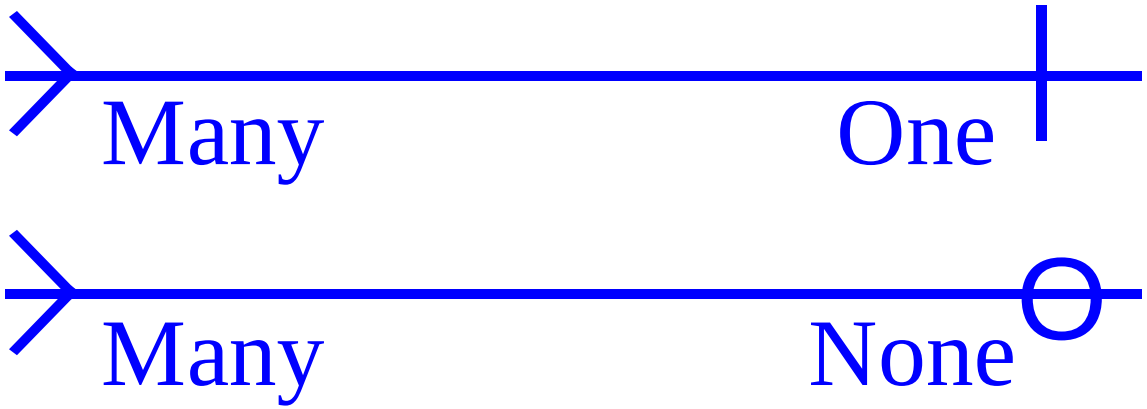
© 2005 Pearson Prentice Hall

Relationships

- Relationships may be:
  - One-to-one.
  - One-to-many.
  - Many-to-many.
- A single vertical line represents one.
- A circle represents zero or none.
- A crows foot represents many.

© 2005 Pearson Prentice Hall

Relationships



© 2005 Pearson Prentice Hall

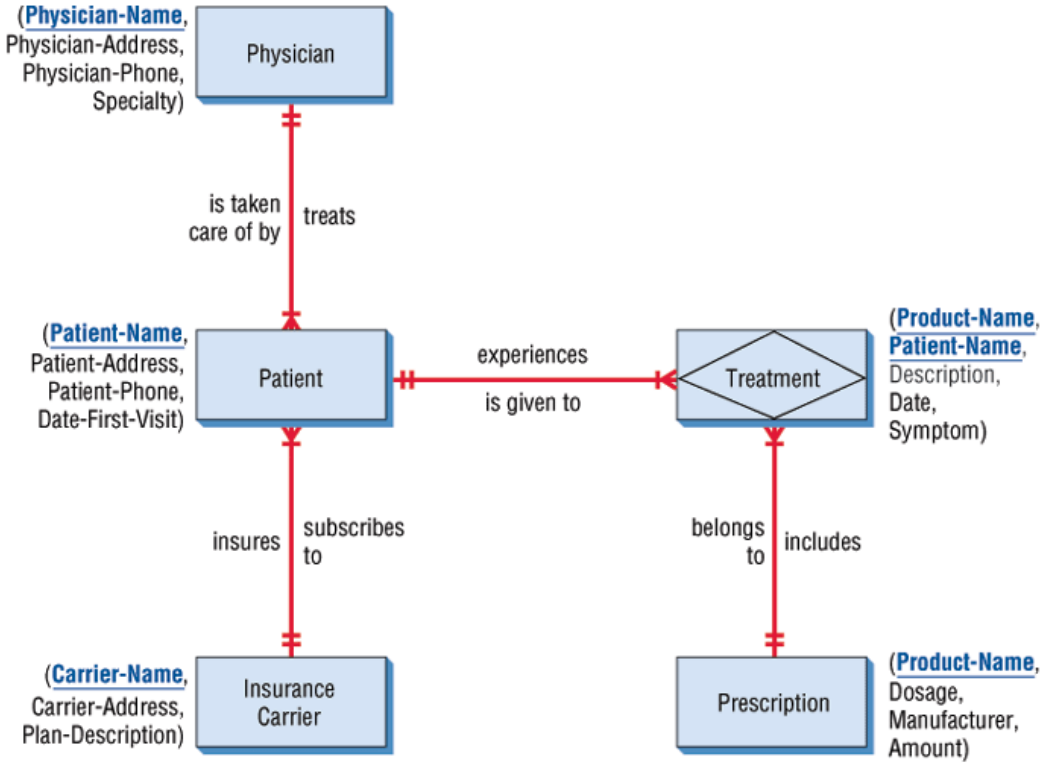
Self-Join

A self-join is when a record has a relationship with another record on the same file.

© 2005 Pearson Prentice Hall

Entity-Relationship Diagram Example

Figure 13.4 The entity-relationship diagram for patient treatment. Attributes can be listed alongside the entities. In each case, the key is underlined.



© 2005 Pearson Prentice Hall

Attributes, Records, and Keys

- Attributes are a characteristic of an entity, sometimes called a data item.
- Records are a collection of data items that have something in common.
- Keys are data items in a record used to identify the record.

© 2005 Pearson Prentice Hall

Key Types

- Key types are:
  - Primary key, unique for the record.
  - Secondary key, a key which may not be unique, used to select a group of records.
  - Concatenated key, a combination of two or more data items for the key.
  - Foreign key, a data item in one record that is the key of another record.

© 2005 Pearson Prentice Hall

Files

- A file contains groups of records used to provide information for operations, planning, management, and decision making.
- Files can be used for storing data for an indefinite period of time, or they can be used to store data temporarily for a specific purpose.

© 2005 Pearson Prentice Hall

File Types

Types of files available are:

- Master file.
- Table file.
- Transaction file.
- Work file.
- Report file.

© 2005 Pearson Prentice Hall

Master and Transaction Files

- Master files
  - Have large records
  - Contain all pertinent information about an entity
- Transaction records
  - Are short records
  - Contain information used to update master files

© 2005 Pearson Prentice Hall

File Organization

- The different organizational structures for file design are:
  - Sequential organization.
  - Linked lists.
  - Hashed file organization.

© 2005 Pearson Prentice Hall

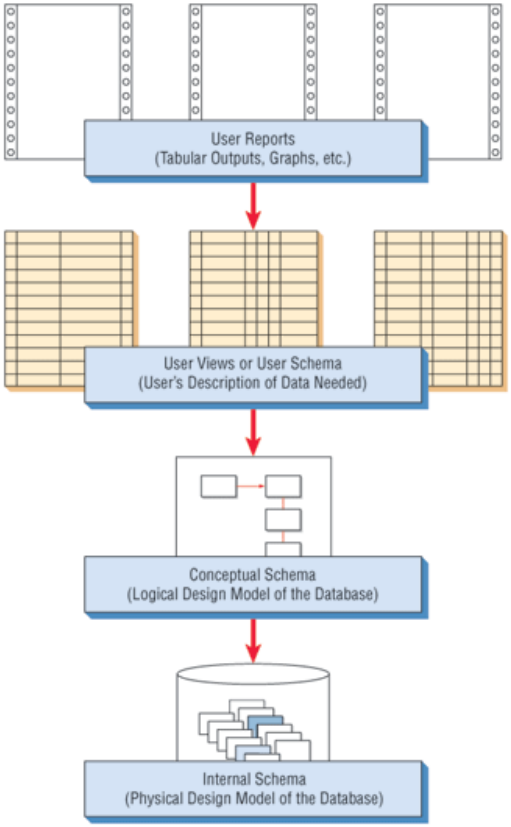
Databases

- A database is intended to be shared by many users.
- There are three structures for storing database files:
  - Relational database structures.
  - Hierarchical database structures (older).
  - Network database structures (older).

© 2005 Pearson Prentice Hall

Logical and Physical Database Design

Figure 13.10 Database design includes synthesizing user reports, user views, and logical and physical designs.



© 2005 Pearson Prentice Hall

Normalization

- Normalization is the transformation of complex user views and data to a set of smaller, stable, and easily maintainable data structures.

© 2005 Pearson Prentice Hall

Normalization (Continued)

- Normalization creates data that are stored only once on a file.
- The exception is key fields.
- The data structures are simpler and more stable.
- The data is more easily maintained.

© 2005 Pearson Prentice Hall

Three Steps of Data Normalization

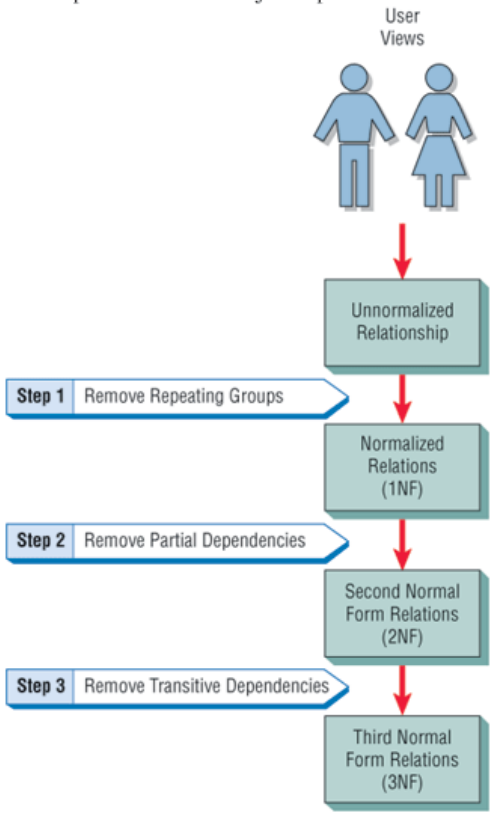
The three steps of data normalization are:

- Remove all repeating groups and identify the primary key.
- Ensure that all nonkey attributes are fully dependent on the primary key.
- Remove any transitive dependencies, attributes that are dependent on other nonkey attributes.

© 2005 Pearson Prentice Hall

Three Steps of Normalization

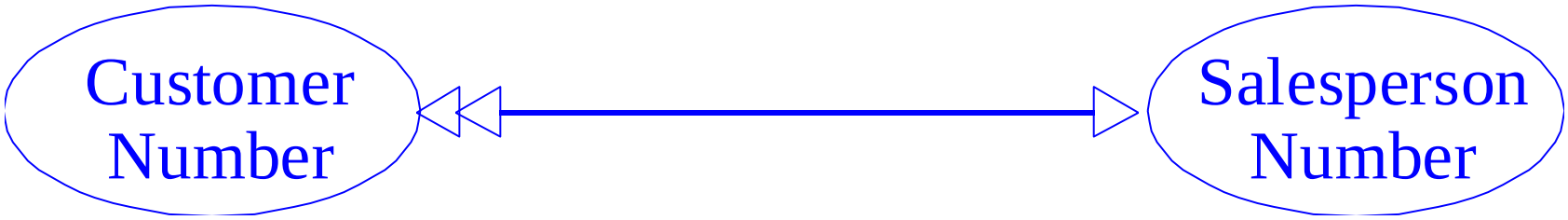
Figure 13.12 Normalization of a relation is accomplished in three major steps.



© 2005 Pearson Prentice Hall

Data Model Diagrams

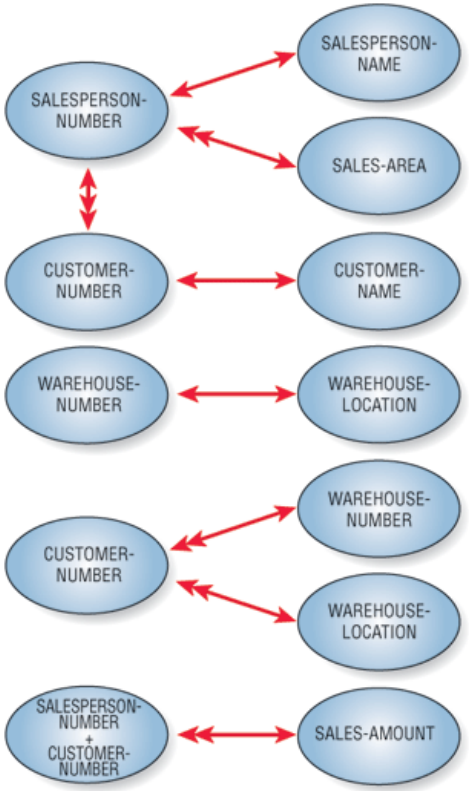
- Data model diagrams are used to show relationships between attributes.
- An oval represents an attribute.
- A single arrow line represents one.
- A double arrow line represents many.



© 2005 Pearson Prentice Hall

Data Model Example

Figure 13.15 Drawing data model diagrams for data associations sometimes helps analysts appreciate the complexity of data storage.



© 2005 Pearson Prentice Hall

First Normal Form (1NF)

- Remove any repeating groups.
- All repeating groups are moved into a new table.
- Foreign keys are used to link the tables.
- When a relation contains no repeating groups, it is in the first normal form.

© 2005 Pearson Prentice Hall

Second Normal Form (2NF)

- Remove any partial dependencies.
- A partial dependency is when the data are only dependent on a part of a key field.
- A relation is created for the data that are only dependent on part of the key and another for data that are dependent on both parts.

© 2005 Pearson Prentice Hall

Third Normal Form (3NF)

- Remove any transitive dependencies.
- A transitive dependency is when a relation contains data that are not part of the entity.
- The problem with transitive dependencies is updating the data.
- A single data item may be present on many records.

© 2005 Pearson Prentice Hall

Entity-Relationship Diagram and Record Keys

The entity-relationship diagram may be used to determine record keys.

- When the relationship is one-to-many, the primary key of the file at the one end of the relationship should be contained as a foreign key on the file at the many end of the relationship.
- A many-to-many relationship should be divided into two one-to-many relationships with an associative entity in the middle.

© 2005 Pearson Prentice Hall

Guidelines for Creating Master Files or Database Relations

Guidelines for creating master files or database relations are:

- Each separate entity should have it's own master file or database relation.
- A specific, nonkey data field should exist on only one master file or relation.
- Each master file or relation should have programs to create, read, update, and delete records.

© 2005 Pearson Prentice Hall

Integrity Constraints

There are three integrity constraints that help to ensure that the database contains accurate data:

- Entity integrity constraints, which govern the composition of primary keys.
- Referential integrity, which governs the denature of records in a one-to-many relationship.
- Domain integrity.

© 2005 Pearson Prentice Hall

Entity Integrity

Entity integrity constraints are rules for primary keys:

- The primary key cannot have a null value.
- If the primary key is a composite key, none of the fields in the key can contain a null value.

© 2005 Pearson Prentice Hall

Referential Integrity

- Referential integrity governs the denature of records in a one-to-many relationship.
- Referential integrity means that all foreign keys in one table (the child table) must have a matching record in the parent table.

© 2005 Pearson Prentice Hall

Referential Integrity (Continued)

Referential integrity includes:

- You cannot add a record without a matching foreign key record.
- You cannot change a primary key that has matching child table records.
  - A child table has a foreign key for a different record.
- You cannot delete a record that has child records.

© 2005 Pearson Prentice Hall

Referential Integrity

- A restricted database updates or deletes a key only if there are no matching child records.
- A cascaded database will delete or update all child records when a parent record is deleted or changed.
- The parent triggers the changes.

© 2005 Pearson Prentice Hall

Domain Integrity

- Domain integrity defines rules that ensure that only valid data are stored on database records
  - Domain integrity has two forms:
    - Check constraints, which are defined at the table level.
    - Rules, which are defined as separate objects and may be used within a number of fields.

© 2005 Pearson Prentice Hall

Retrieving and Presenting Database Data

- The guidelines to retrieve and present data are:
  - Choose a relation from the database.
  - Join two relations together.
  - Project columns from the relation.
  - Select rows from the relation.
  - Derive new attributes.
  - Index or sort rows.
  - Calculate totals and performance measures.
  - Present data.

© 2005 Pearson Prentice Hall

Denormalization

Denormalization is the process of taking the logical data model and transforming it into an efficient physical model.

© 2005 Pearson Prentice Hall

Data Warehouses

Data warehouses are used to organize information for quick and effective queries.

© 2005 Pearson Prentice Hall

Data Warehouses and Database Differences

- In the data warehouse, data are organized around major subjects.
- Data in the warehouse are stored as summarized rather than detailed raw data.
- Data in the data warehouse cover a much longer time frame than in a traditional transaction-oriented database.

© 2005 Pearson Prentice Hall

Data Warehouses and Database Differences (Continued)

- Data warehouses are organized for fast queries.
- Data warehouses are usually optimized for answering complex queries, known as OLAP.
- Data warehouses allow for easy access via data-mining software called software.

© 2005 Pearson Prentice Hall

Data Warehouses and Database Differences (Continued)

- Data warehouses include multiple databases that have been processed so that data are uniformly defined, containing what is referred to as "clean" data.
- Data warehouses usually contain data from outside sources.

© 2005 Pearson Prentice Hall

Online Analytic Processing (OLAP)

- Online analytic processing (OLAP) is meant to answer decision makers' complex questions by defining a multidimensional database.
- Data mining, or knowledge data discovery (KDD), is the process of identifying patterns that a human is incapable of detecting.

© 2005 Pearson Prentice Hall

Data Mining Decision Aids

Data mining has a number of decision aids available, including:

- Statistical analysis.
- Decision trees.
- Neural networks.
- Intelligent agents.
- Fuzzy logic.
- Data visualization.

© 2005 Pearson Prentice Hall

Data Mining Patterns

Data mining patterns that decision makers try to identify include:

- Associations, patterns that occur together.
- Sequences, patterns of actions that take place over a period of time.
- Clustering, patterns that develop among groups of people.
- Trends, the patterns that are noticed over a period of time.

© 2005 Pearson Prentice Hall

Web Based Databases and XML

- Web-based databases are used for sharing data.
- Extensible markup language (XML) is used to define data used primarily for business data exchange over the Web.

© 2005 Pearson Prentice Hall