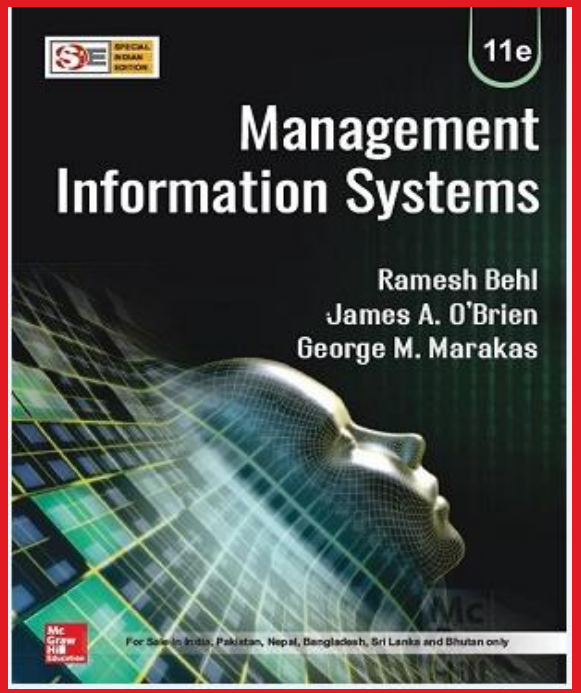
Because learning changes everything.®

Management Information Systems

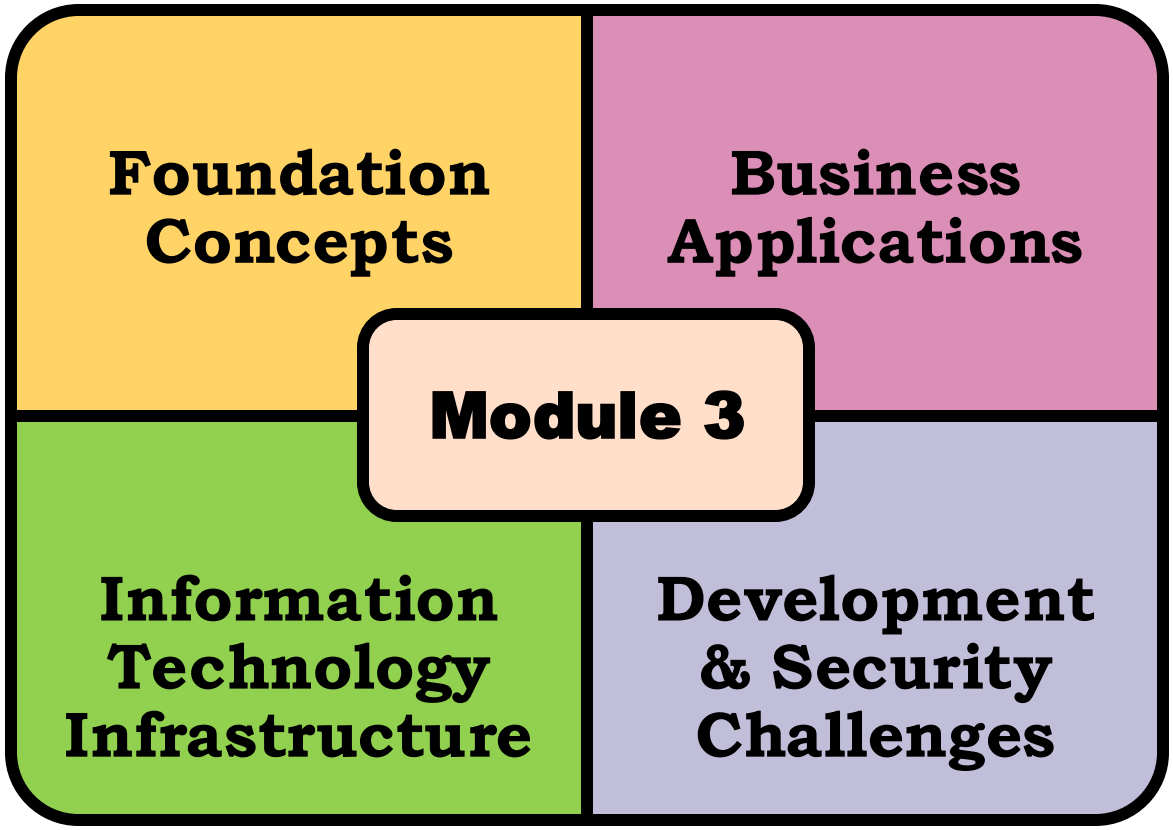
Eleventh Edition



Copyright © 2019 McGraw Hill Education, All Rights Reserved.

PROPRIETARY MATERIAL © 2019 The McGraw Hill Education, Inc. All rights reserved. No part of this PowerPoint slide may be displayed, reproduced or distributed in any form or by any means, without the prior written permission of the publisher, or used beyond the limited distribution to teachers and educators permitted by McGraw Hill for their individual course preparation. If you are a student using this PowerPoint slide, you are using it without permission.

Chapter 8 : Data Resource Management


Foundation Concepts
Business Applications
Module 3
Information Technology Infrastructure
Development & Security Challenges

Learning Objectives

Explain the business value of implementing data resource management processes and technologies in an organization.

Outline the advantages of a database management approach to managing the data resources of a business, compared with a file processing approach.

Understand how database management software helps business professionals and supports the operations and management of a business.

Understand types of databases, fundamental database structures and database development methods.

Understand data warehouse and data mining concepts and its applications to big data.

RWC 1: Data-Driven Crime Fighting

Information-led policing

- Era began in 1990 in New York City

Crime mapping

- Incident reporting combined with GIS

Predictive policing

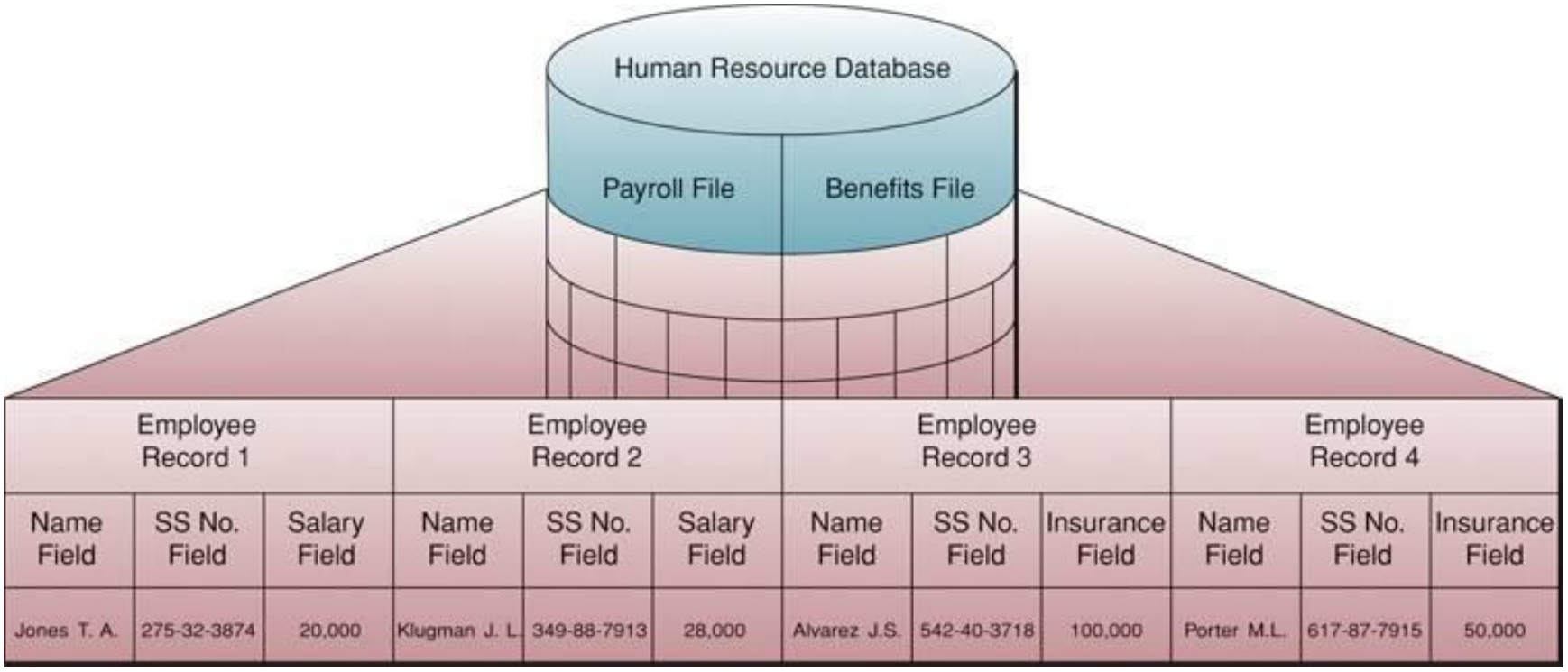
- Predict where and when crimes will occur

Pertinent questions

- How long is data kept?
- With whom is it shared?

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Logical Data Elements



Employee Record 1			Employee Record 2			Employee Record 3			Employee Record 4		
Name Field	SS No. Field	Salary Field	Name Field	SS No. Field	Salary Field	Name Field	SS No. Field	Insurance Field	Name Field	SS No. Field	Insurance Field
Jones T. A.	275-32-3874	20,000	Klugman J. L.	349-88-7913	28,000	Alvarez J.S.	542-40-3718	100,000	Porter M.L.	617-87-7915	50,000

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Logical Data Elements

File or table

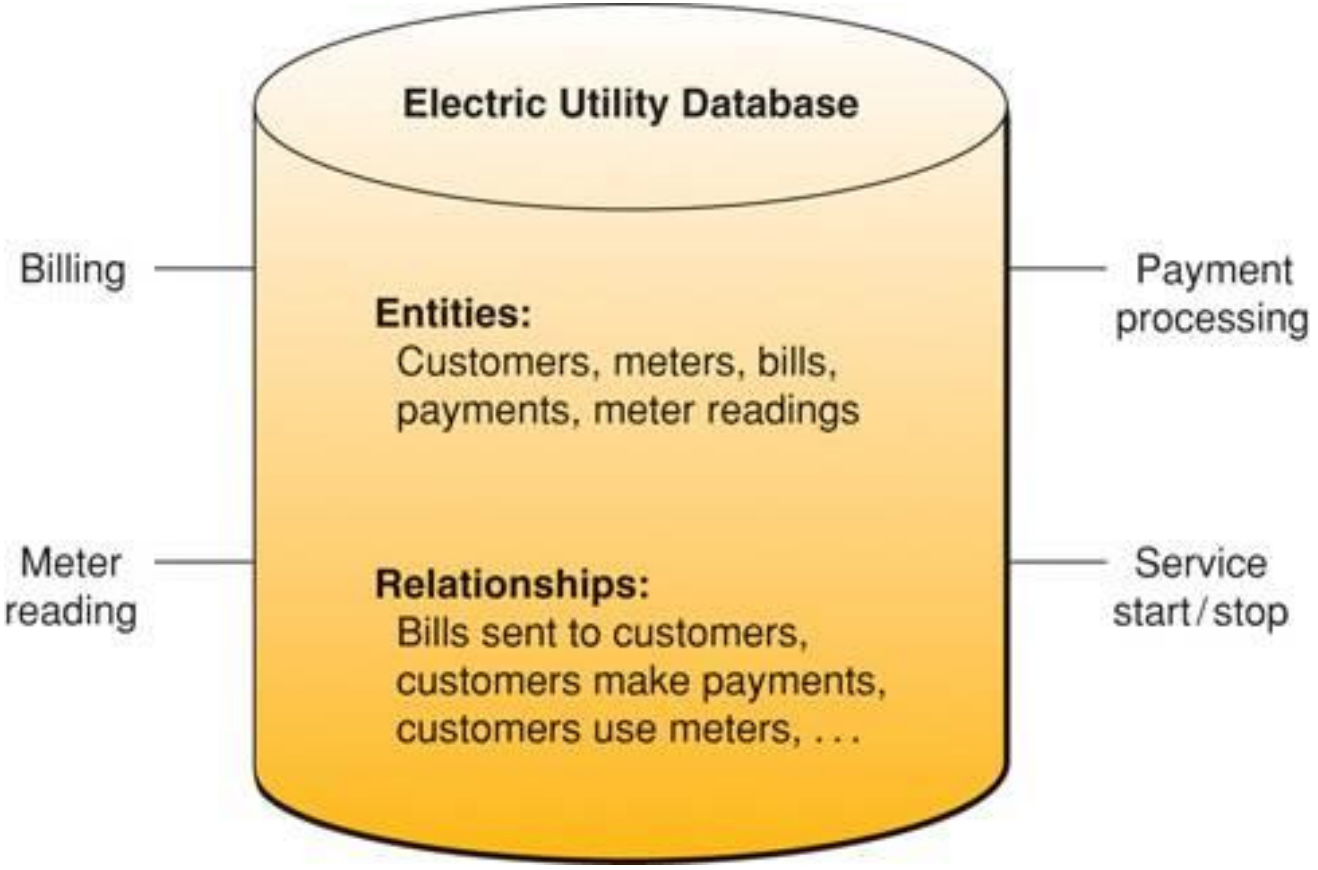
- A group of related records
- Master file
- Transaction file
- History file
- Archival file

Database

- An integrated collection of logically related data elements
  - World's largest database?

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Electric Utility Database



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Structures

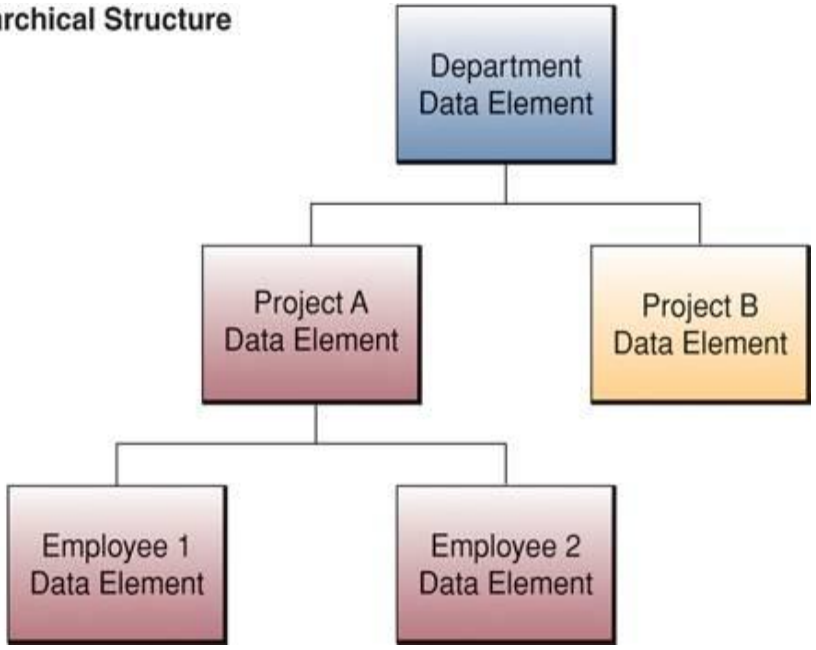
Common database structures…

- Hierarchical
- Network
- Relational
- Object-oriented
- Multi-dimensional

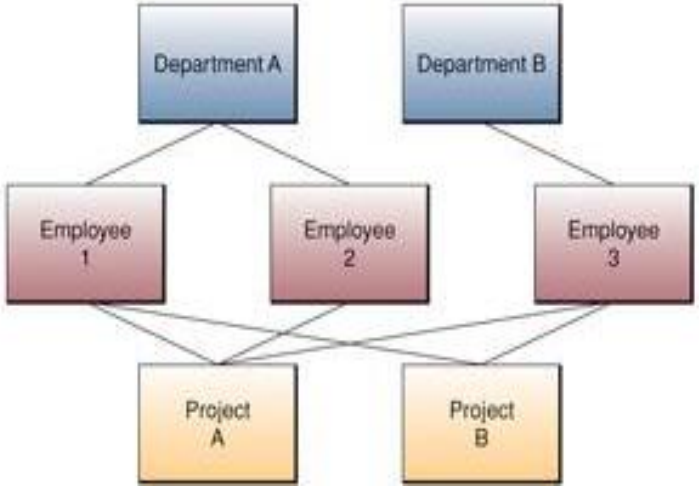
Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Structures

Hierarchical Structure



Network Structure

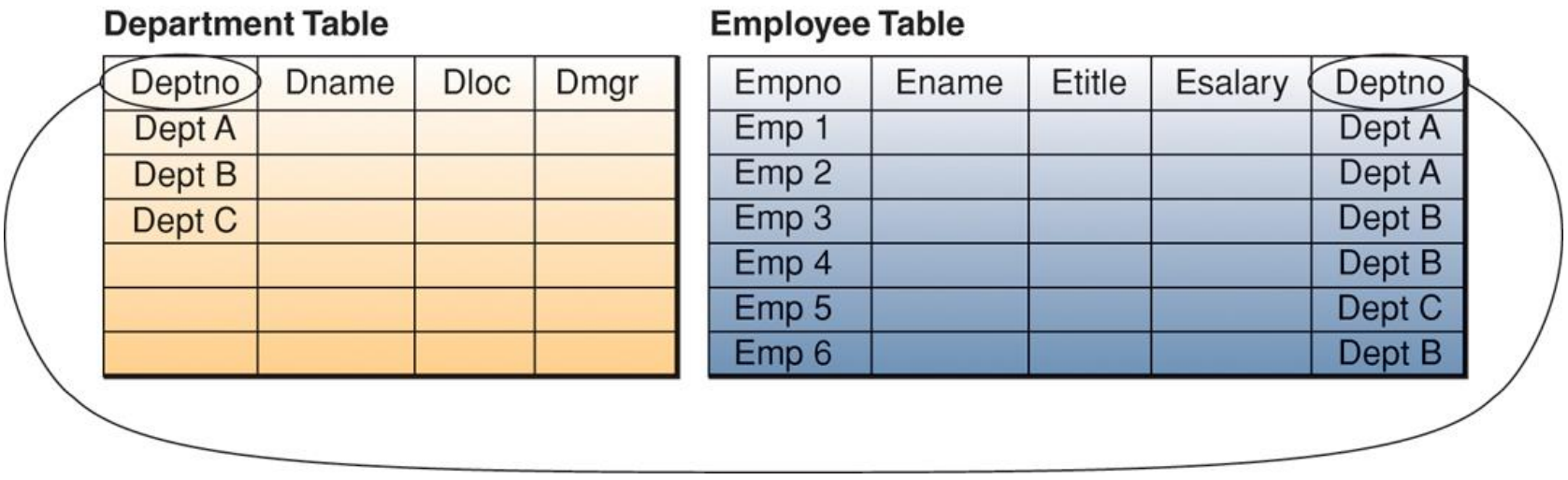


Copyright © 2019 by McGraw Hill Education. All rights reserved.

Relational Structure

Most widely used structure

- Data elements are stored in tables
- Row represents a record; column is a field
- Can relate data in one file with data in another, if both files share a common data element



Department Table

Deptno	Dname	Dloc	Dmgr
Dept A			
Dept B			
Dept C			

Employee Table

Empno	Ename	Etitle	Esalary	Deptno
Emp 1				Dept A
Emp 2				Dept A
Emp 3				Dept B
Emp 4				Dept B
Emp 5				Dept C
Emp 6				Dept B

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Relational Operations

Select

- Create a subset of records that meet a stated criterion
  - Example: employees earning more than $30,000

Join

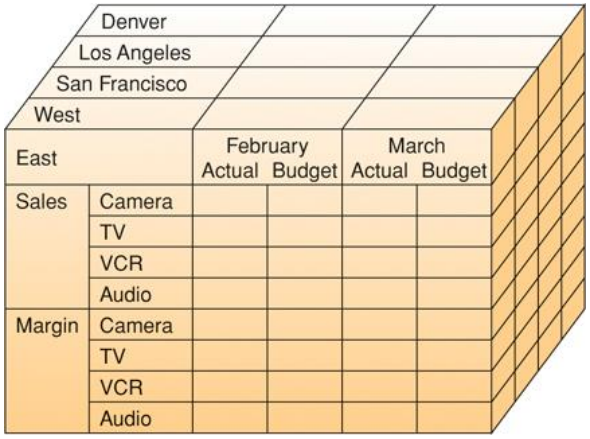
- Combine two or more tables temporarily
- Looks like one big table

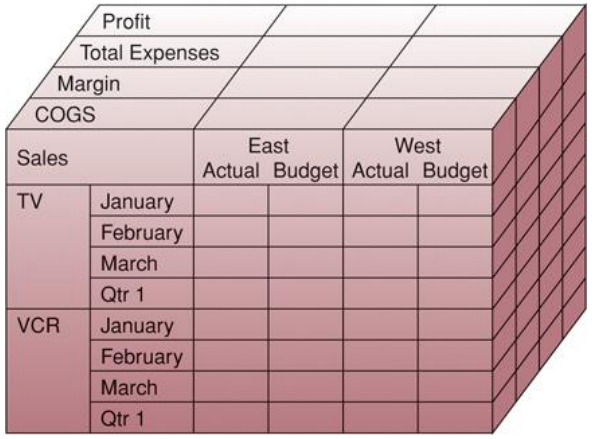
Project

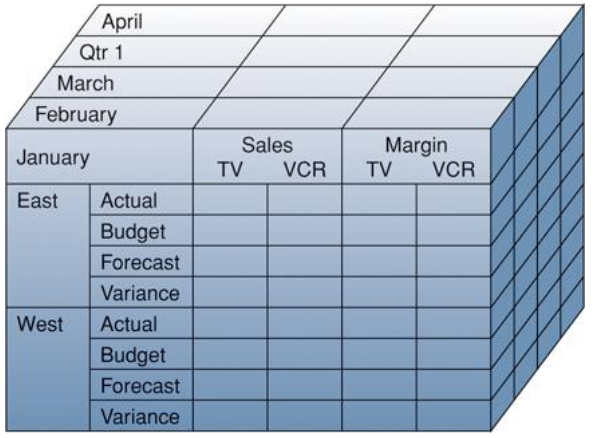
- Create a subset of columns in a table

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Multidimensional Model

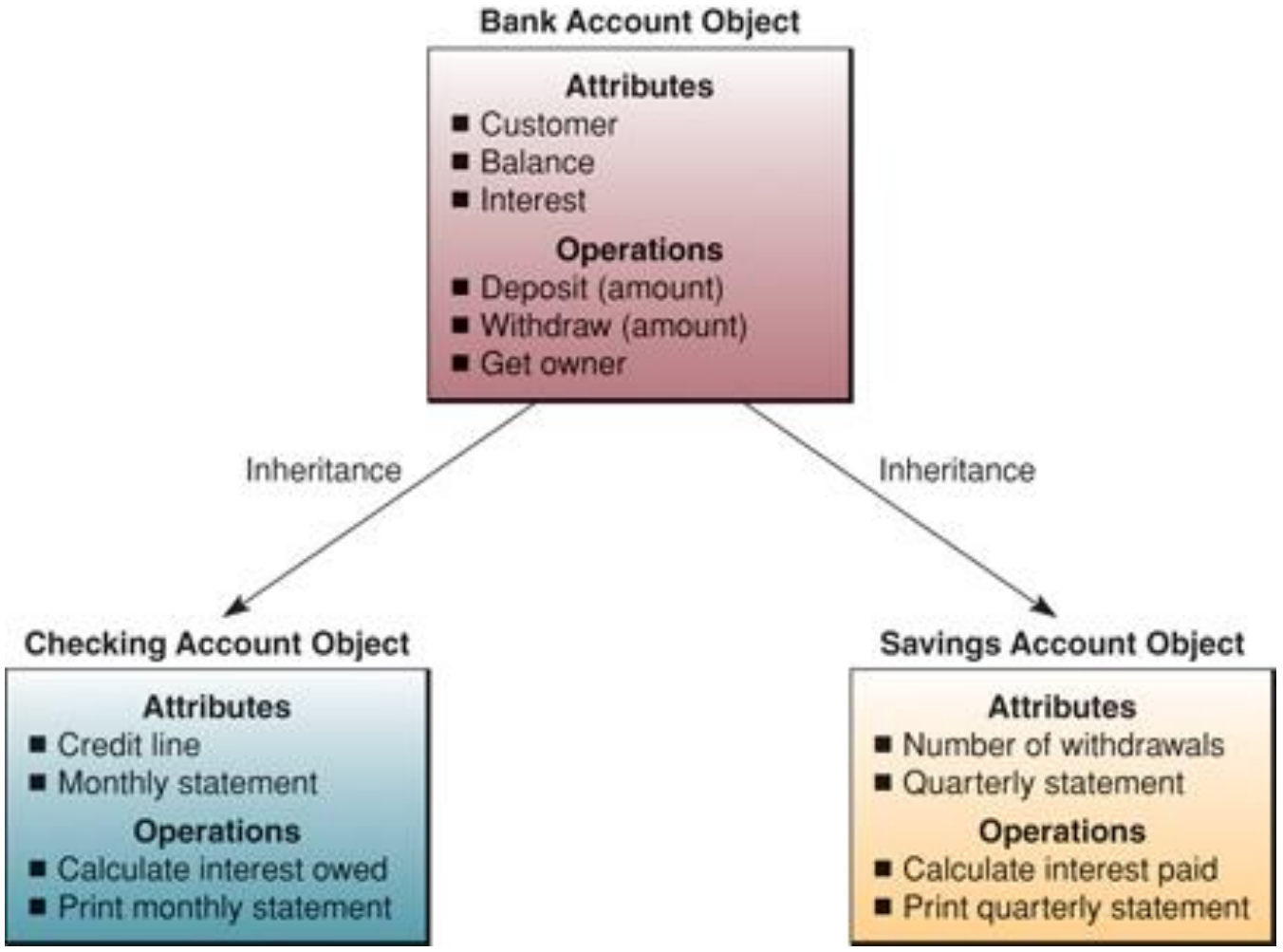






Copyright © 2019 by McGraw Hill Education. All rights reserved.

Object-Oriented Structure



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Evaluation of Database Structures

Hierarchical

- Works for structured, routine transactions
- Can’t handle many-to-many relationships

Network

- More flexible than hierarchical
- Unable to handle ad hoc requests

Relational

- Easily responds to ad hoc requests
- Easier to work with and maintain
- Not as efficient/quick as hierarchical or network

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Development

Database Administrator (DBA)

- Enterprise database development
- Improves integrity and security
- Data Definition Language (DDL)
  - Data contents, relationships, and structure
- Specifications
  - Data dictionary
  - Metadata repository

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Data Dictionary

A data dictionary

- Contains data about data (metadata)
- Specialized software manages data definitions

Contains information on…

- Names, types and descriptions of data
- Relationships
- Requirements for access and use
- Maintenance
- Security

Copyright © 2019 by McGraw Hill Education. All rights reserved.

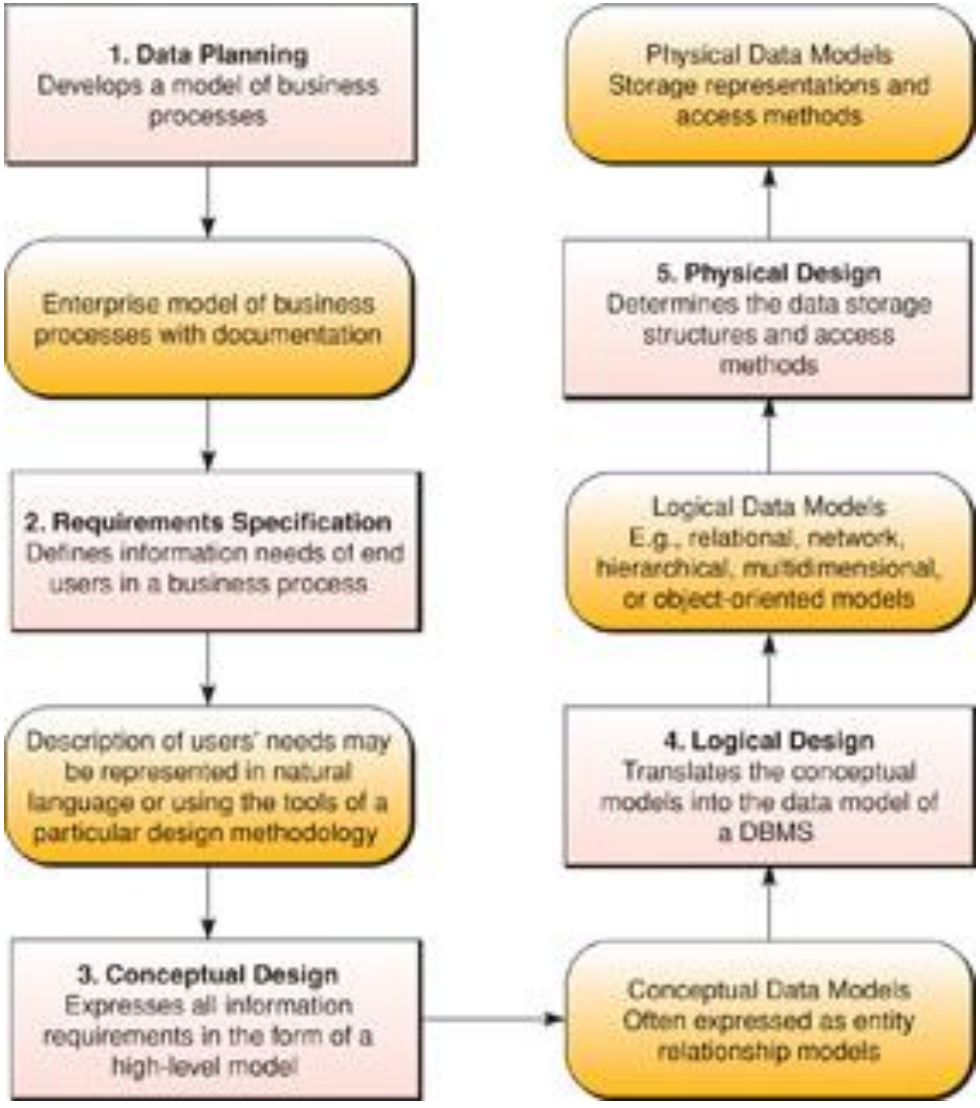
Data Planning and Database Design

Database development is top-down process

- Develop an enterprise model
- Define the information needs of end users
- Identify the key data elements

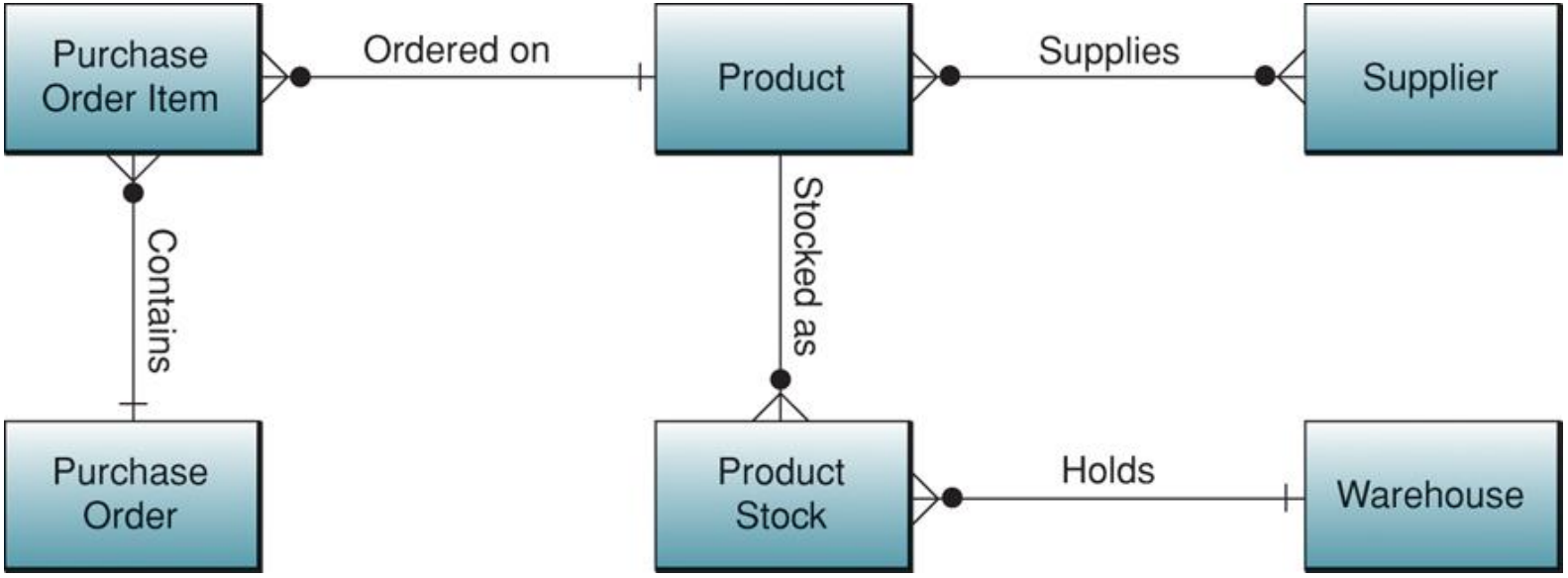
Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Development



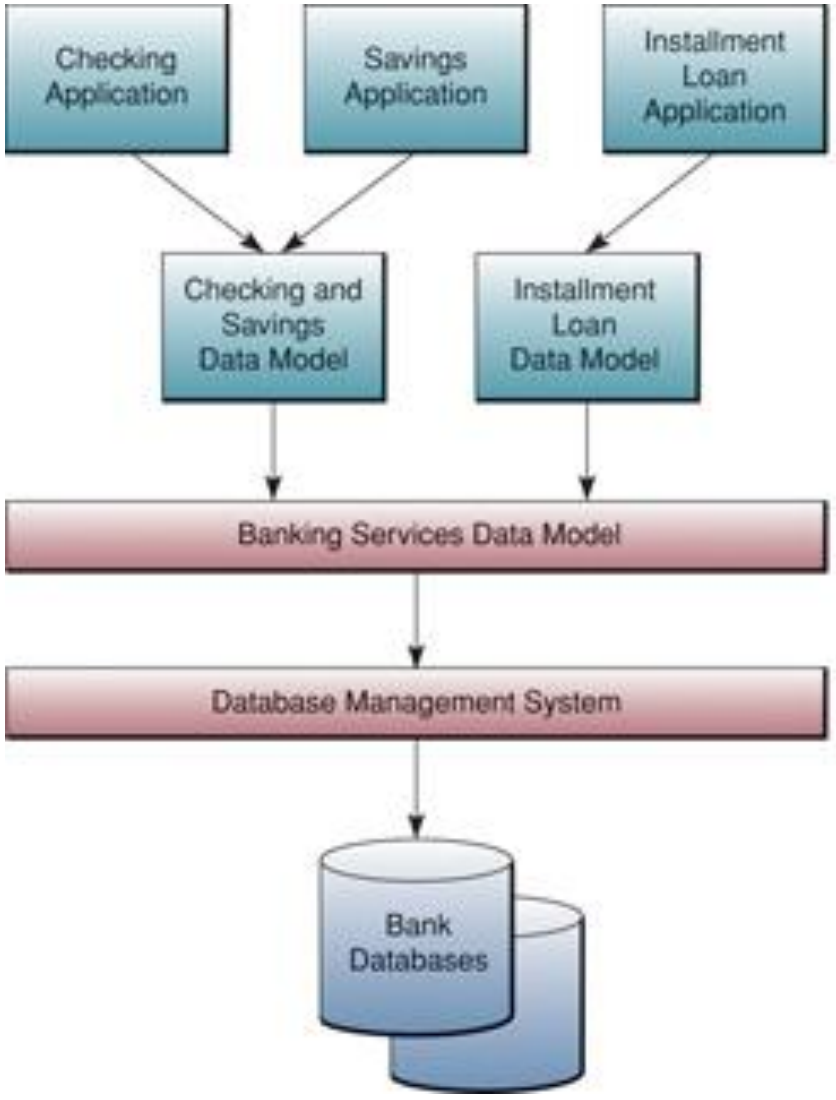
Copyright © 2019 by McGraw Hill Education. All rights reserved.

Entity Relationship Diagram



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Logical and Physical Database Views



Logical User Views
Data elements and relationships (the subschemas) needed for checking, savings, or installment loan processing

Data elements and relationships (the schema) needed for the support of all bank services

Software Interface
The DBMS provides access to the bank's databases

Physical Data Views
Organization and location of data on the storage media

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Data Resource Management

Managerial activity

- Uses data management, data warehousing, and other IS technologies
- Manages data for business stakeholders

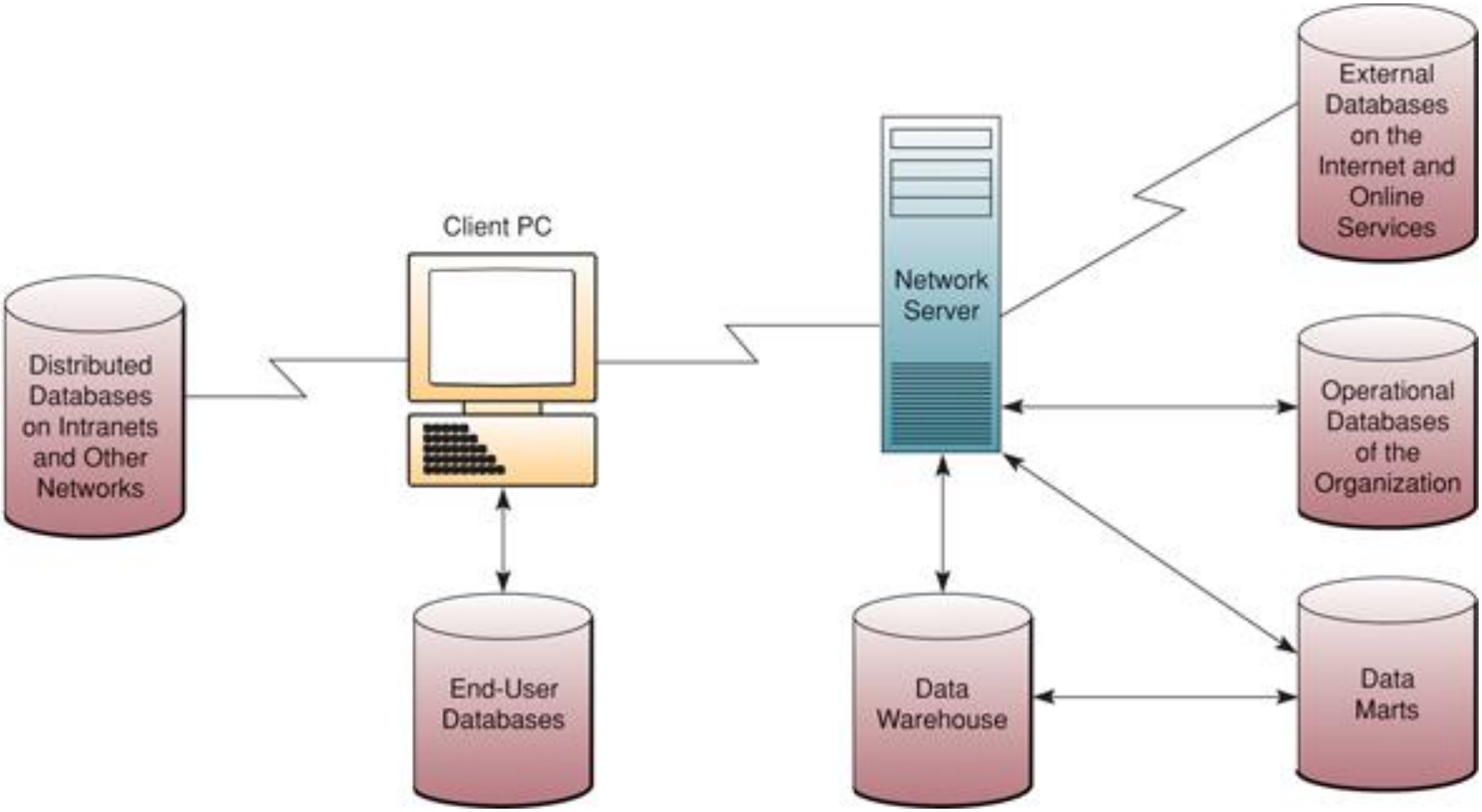
Copyright © 2019 by McGraw Hill Education. All rights reserved.

RWC 2: Medical IT Is Getting Personal

- Duke University Health System
  - Personalized approach to treating patients
  - Analytics engine on top of clinical repository
  - Decision support tools pick best treatment
- Beth Israel Deaconess Medical Center
  - Catch details that elude a doctor
  - Tests personalized for patient
- National Cancer Institute
  - Blend data warehouse with Web collection tools
  - More patient data for research

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Types of Databases



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Operational Databases

- also known as transactional databases,
- designed to handle day-to-day operations of an organization.
- optimized for fast read and write operations.
- commonly used in applications such as
  - e-commerce,
  - banking,
  - inventory management, etc.
- Examples include Oracle Database, MySQL, PostgreSQL.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Distributed Database

- designed to store data across multiple nodes or servers.
- provide scalability, fault tolerance, and high availability.
- data is partitioned and distributed across different nodes, allowing for parallel processing and improved performance.
- commonly used in large-scale applications and distributed systems.
- Examples include Apache Cassandra, Apache HBase, Riak.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Distributed Database

Challenges

- **Data Consistency:** Ensuring synchronized and consistent data across multiple nodes.
- **Data Partitioning and Distribution:** Dividing and distributing data across nodes for improved performance and scalability.
- **Network Communication and Latency:** Dealing with delays and disruptions in network communication between distributed nodes.
- **Distributed Query Processing:** Optimizing query performance across distributed nodes.
- **Scalability and Load Balancing:** Efficiently scaling the system and distributing workload evenly among nodes.
- **Failure Handling and Fault Tolerance:** Handling node failures and ensuring system resilience.
- **Operational Complexity**: Managing the complexity associated with configuring, monitoring, and administering distributed databases.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

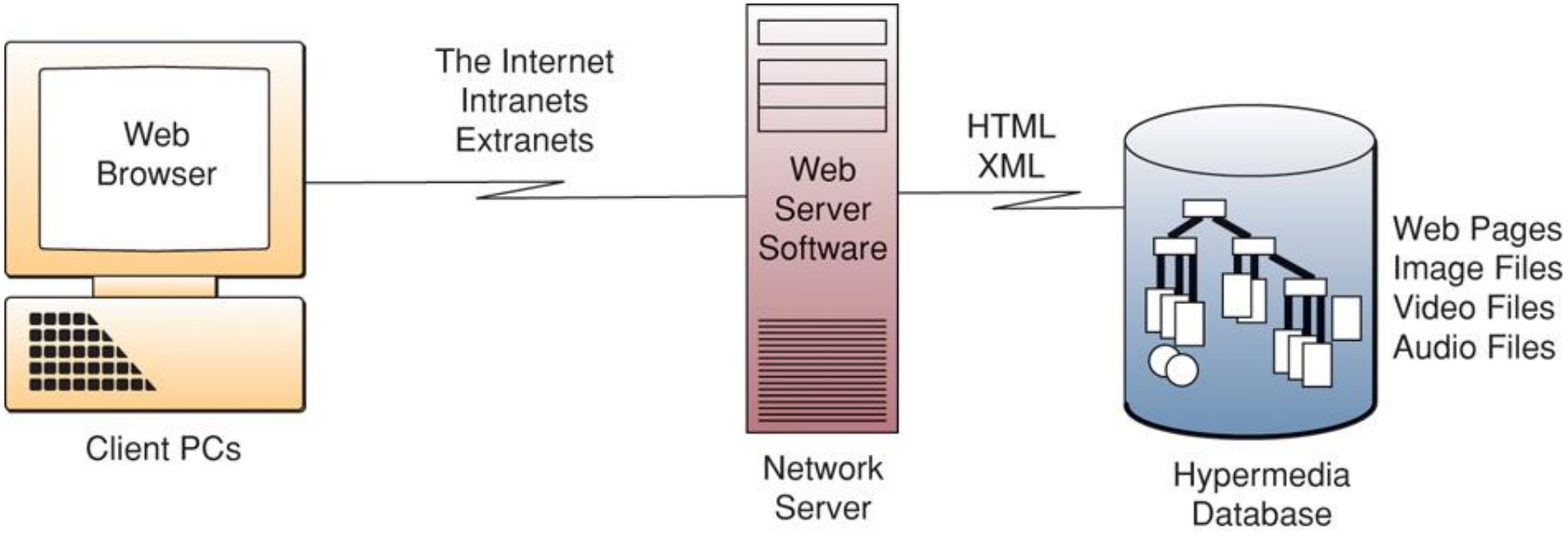
External Database

- also known as an external data source or external data store,
- is external or separate from the primary application or system that is using it
- The primary application or system interacts with the external database by making requests to fetch or update data through the provided interfaces or APIs.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Hyperlink Databases

- support the storage and retrieval of hypermedia content.
- provide functionality to store and manage text, images, audio, video, and associated hyperlinks within the database.



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Big Data

Big Data is variety of voluminous data that is changing very fast. The three important characteristics of Big Data are:

- **Variety:** Big data comprises of data of all types and all formats. Some are structured in the form of traditional relational databases and some are unstructured in the for of emails, chats, SMSs, videos, audios etc.
- **Volume**: Big data collects data from all sources including internal and external and process the same to find something meaningful.
- **Velocity:** Big Data capture and processes the data in unprecedented speed.

NoSQL database is designed to handle large variety of data models.

Hadoop Apache Hadoop is an open-source framework designed for the distributed storage and processing of large datasets using a simple programming model.

MapReduceMapReduce is a programming model and processing engine for distributed computing that processes large datasets in parallel across a distributed cluster of processors or nodes.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Data Warehouse

Data Warehouse: It is a subject-oriented, integrated, time-varient and non-volatile collection of data.

Data Mart: It is a subset of a data warehouse

Data Mining: It is the process of discovering hidden patterns/knowledge from the data repository.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Characteristics of Data Warehouse

Subject-Oriented:

A data warehouse is organized around specific subject areas relevant to the business, such as sales, finance, or customer relations. It provides a consolidated view of data related to these subjects.

Integrated:

Data from diverse sources, including operational databases, spreadsheets, and external data, is integrated and transformed into a common format. This ensures consistency and uniformity in the data.

Time-Variant:

Data in a data warehouse is stored and managed over time, allowing for historical analysis. It includes a timeline, enabling users to analyze trends, changes, and performance over different time periods.

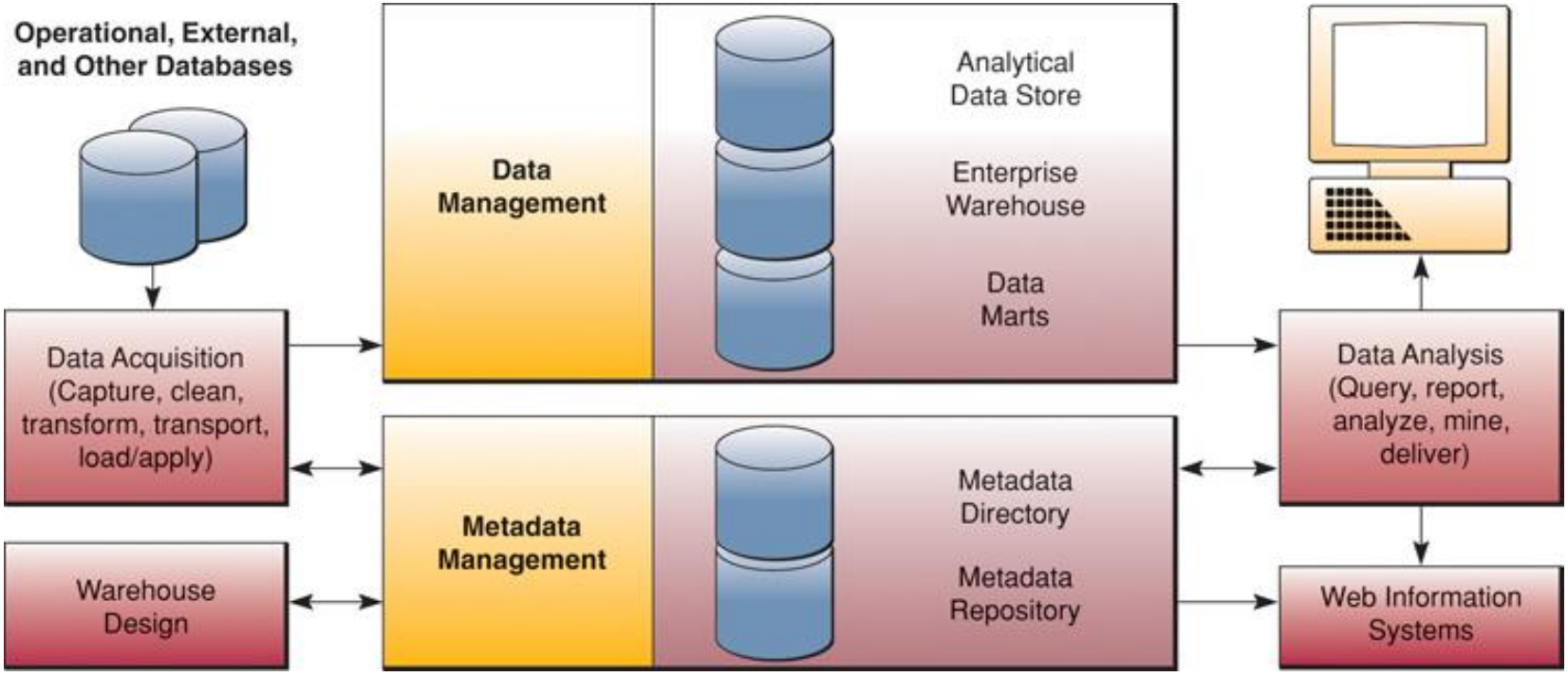
Non-Volatile:

Data in a data warehouse is non-volatile, meaning that once it is loaded into the system, it is not typically subject to frequent changes. This stability ensures a reliable and consistent data source for analysis.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

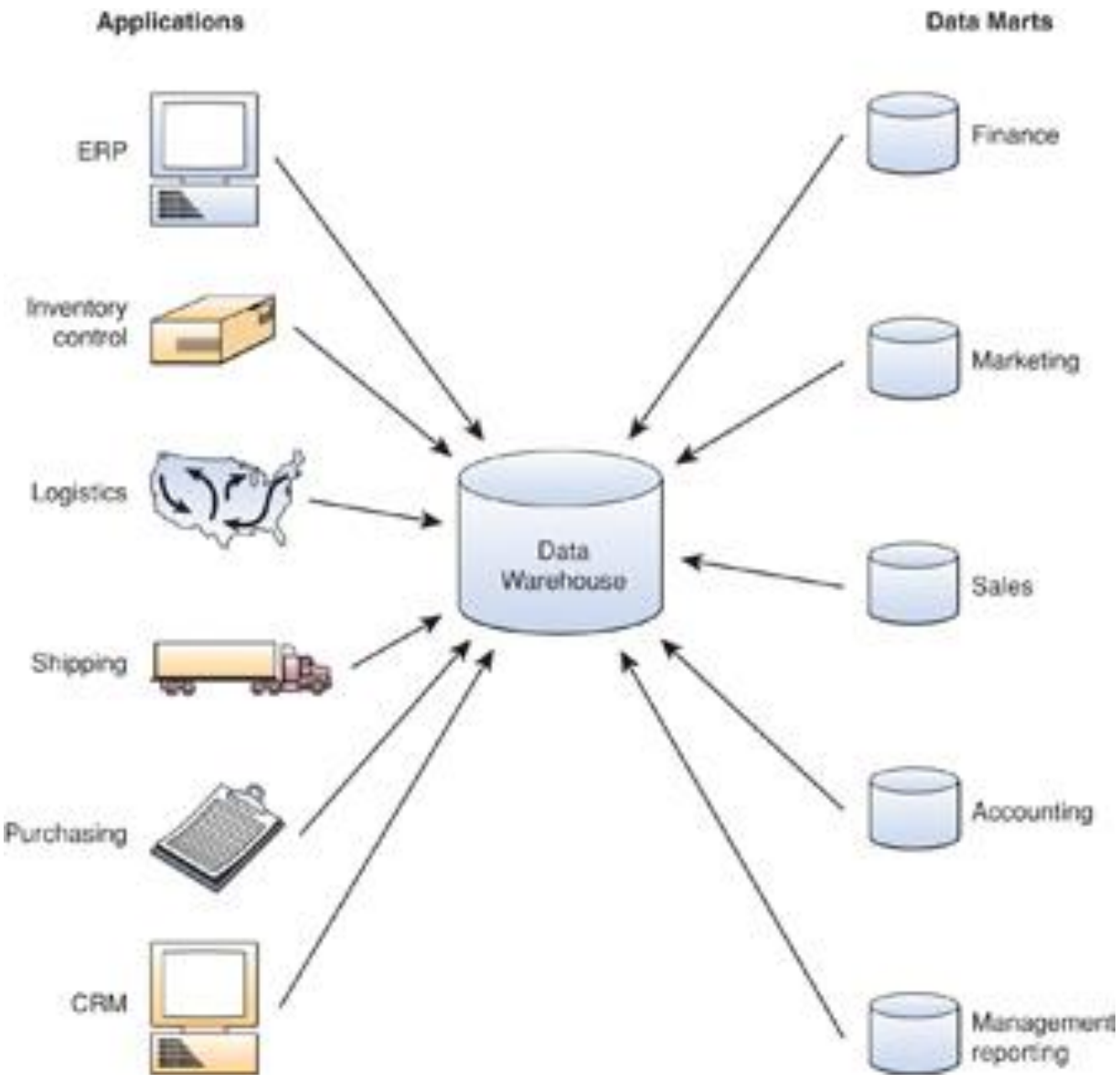
Aspect	Database	Data Warehouse
Purpose	Transactional processing and operations	Analytical processing and decision support
Data Structure	Normalized for minimizing redundancy	Denormalized for analytical efficiency
Data Type	Handles current, transactional data	Stores historical data over time
Query and Reporting	Optimized for simple queries and real-time operations	Optimized for complex queries, aggregations, and reporting
Performance	Optimized for transactional performance	Optimized for analytical performance
Example Technologies	MySQL, PostgreSQL, Oracle Database, MongoDB, Cassandra	Amazon Redshift, Google BigQuery, Snowflake, Teradata
Usage Scenario	Day-to-day operational tasks, real-time data retrieval	Analytical tasks, business intelligence, decision support
Data Focus	Current and frequently updated data	Historical and aggregated data

Data Warehouse Components



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Data Warehouse and Data Marts



Copyright © 2019 by McGraw Hill Education. All rights reserved.

The ETL Process

The ETL (Extract, Transform and Load) process is one of the most important stages of data warehouse.

Extract: It extracts the data from the source systems – operational databases, external databases and ERP systems. Data from different sources is converted into one consolidated data warehouse format, which is then put for transformation exercise.

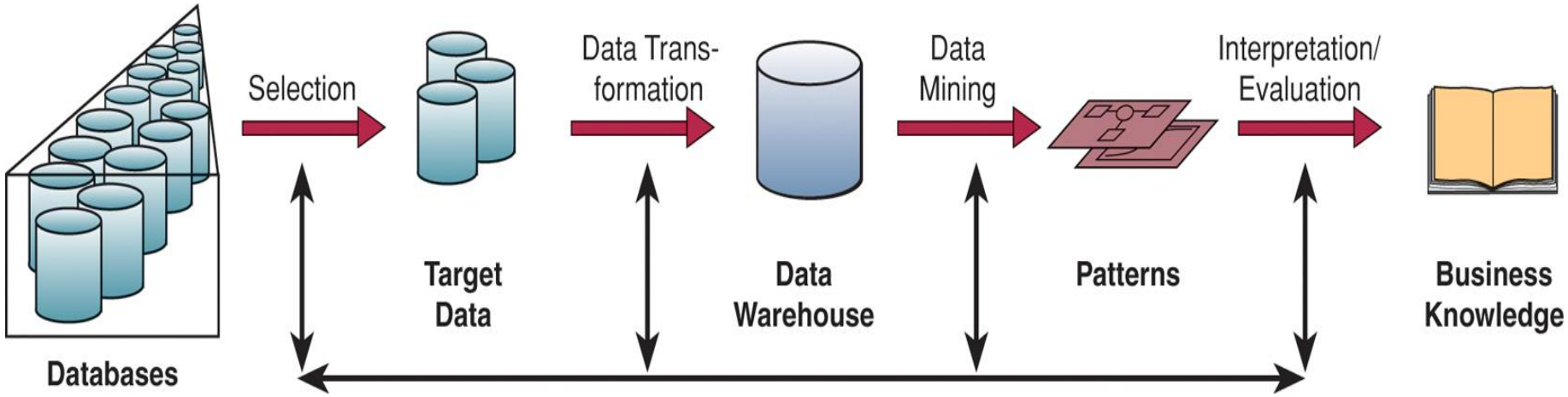
Transform: At this stage following tasks may be performed:

- Cleansing of data
- Filtering of data
- Applying Business rules
- Joining data from multiple sources
- Spliting of data
- Transposing rows and columns
- Aggregations or dis-aggregations

Loading: After applying the transformation rules data is ready to be loaded to the warehouse.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Data Mining



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Why Mine Data?

1. Visualizing Data:
   - **Purpose:** Understand and make sense of large datasets stored in data warehouses or databases.
   - **Importance:** Visualization tools help uncover patterns and insights, providing a comprehensive understanding of the data.
2. Discovering Knowledge:
   - **Purpose:** Uncover hidden relationships, patterns, and correlations within enterprise databases.
   - **Importance:** Knowledge discovery contributes to informed decision-making by extracting valuable insights from the data.
3. Sharing the Story:
   - **Purpose:** Share findings with stakeholders to develop future strategies.
   - **Importance:** Tools for simple and pictorial presentation help communicate insights, enabling informed strategic planning and actions.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Applications of Data Mining

1. Retail Industry:
   - **Market Basket Analysis:** Identifying associations and patterns in customer shopping baskets to optimize product placement, promotions, and inventory management.
   - **Customer Segmentation:** Segmenting customers based on their purchasing behavior to personalize marketing strategies and promotions.
2. Finance and Banking:
   - **Credit Scoring:** Assessing the creditworthiness of individuals or businesses by analyzing historical financial data.
   - **Fraud Detection**: Identifying unusual patterns and anomalies in financial transactions to detect fraudulent activities.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Applications of Data Mining

Marketing and Advertising:

- **Targeted Marketing:** Analyzing customer data to create targeted marketing campaigns and personalized advertisements.
- **Customer Retention**: Identifying factors influencing customer loyalty and designing strategies to retain customers..

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Applications of Data Mining

Manufacturing and Production:

- Quality Control: Analyzing production data to identify patterns that indicate potential quality issues and optimize manufacturing processes.
- Supply Chain Optimization: Analyzing supply chain data to optimize inventory levels, reduce costs, and improve efficiency.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Applications of Data Mining

E-commerce:

- Recommendation Systems: Analyzing customer preferences and behavior to provide personalized product recommendations.
- Dynamic Pricing: Analyzing market conditions and customer behavior to optimize pricing strategies dynamically.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

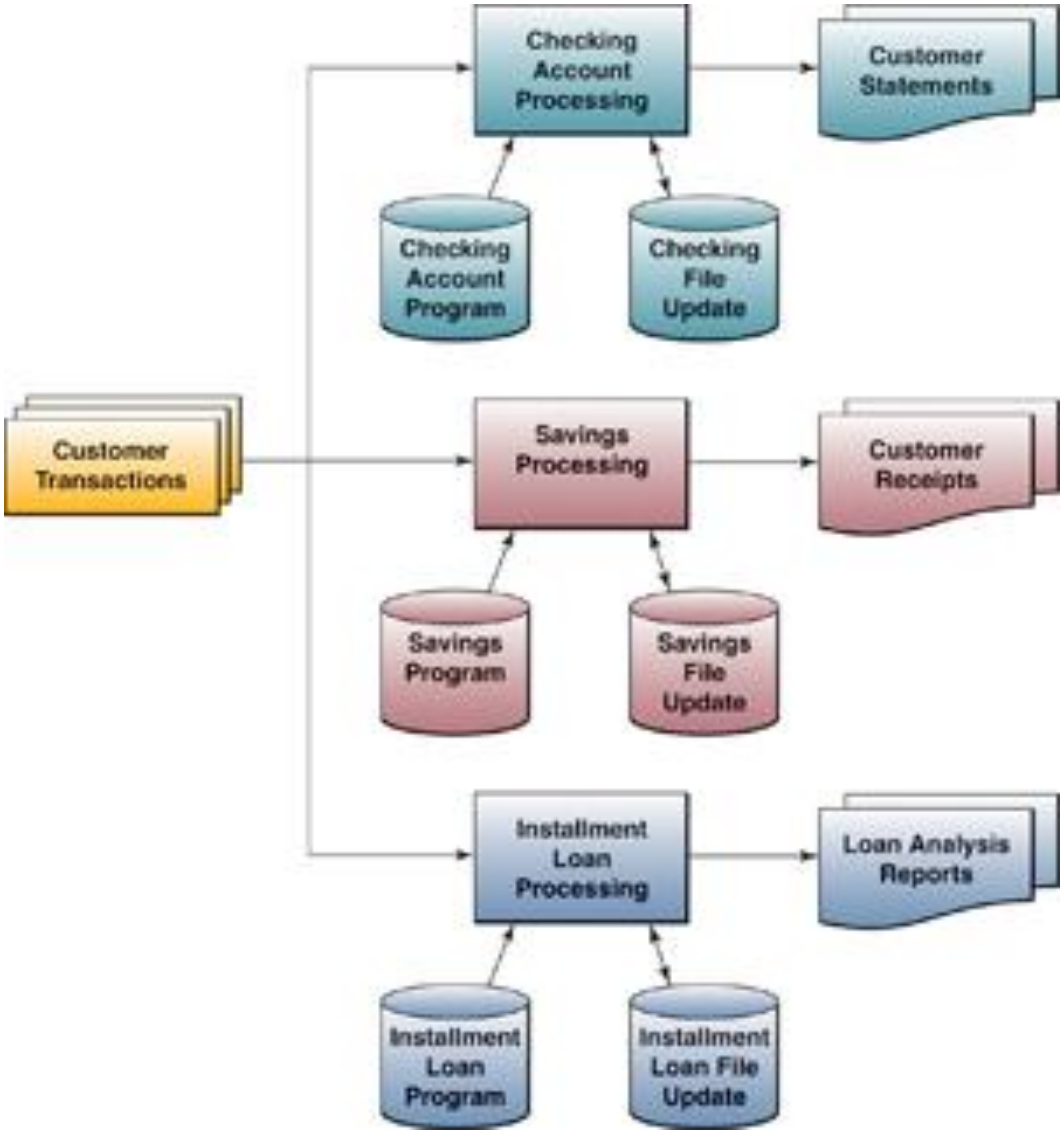
Applications of Data Mining

Transportation and Logistics:

- Route Optimization: Analyzing traffic and logistics data to optimize transportation routes and schedules.
- Predictive Maintenance for Vehicles: Predicting maintenance needs for vehicles based on historical performance data.

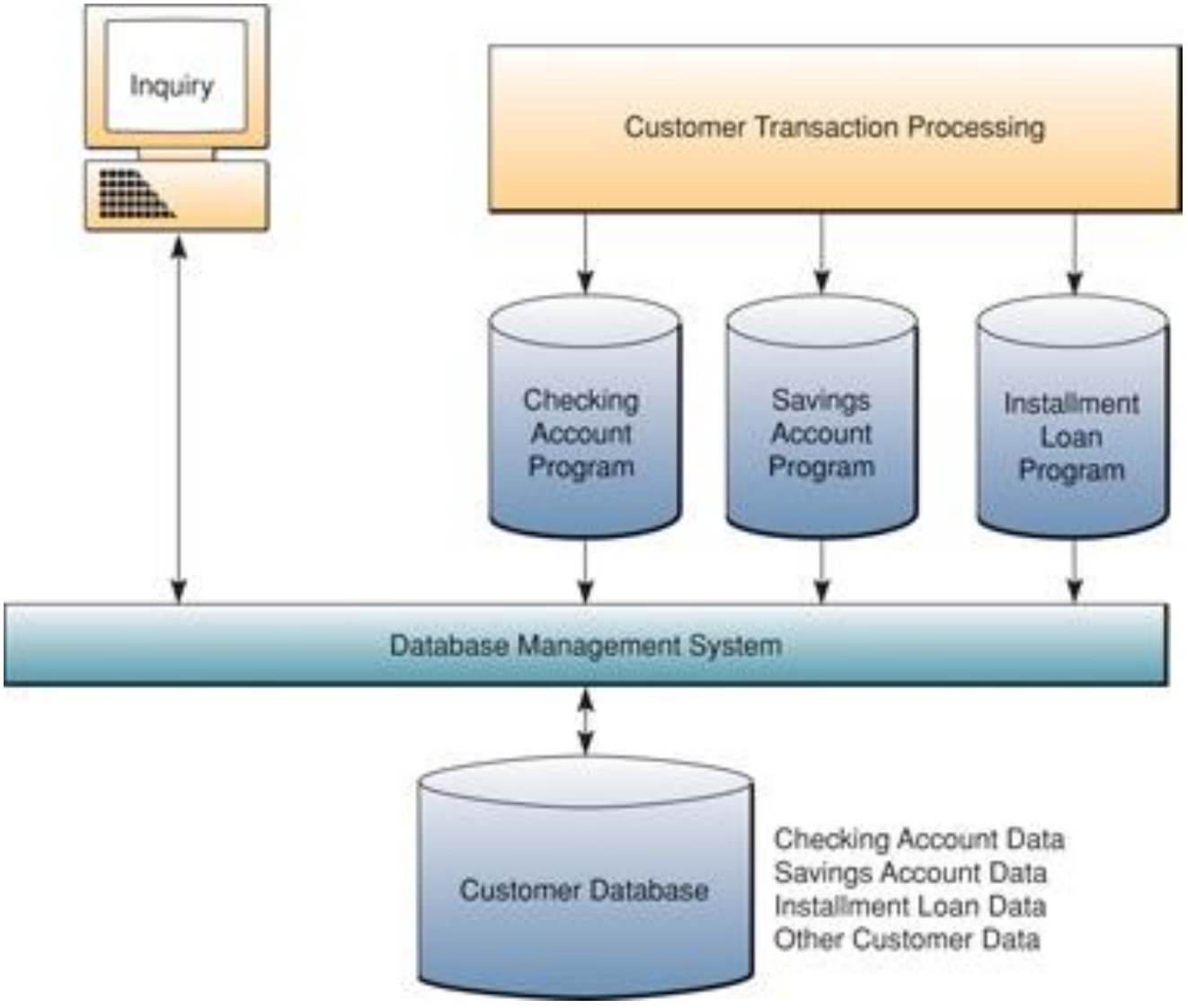
Copyright © 2019 by McGraw Hill Education. All rights reserved.

Traditional File Processing



Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Management Approach



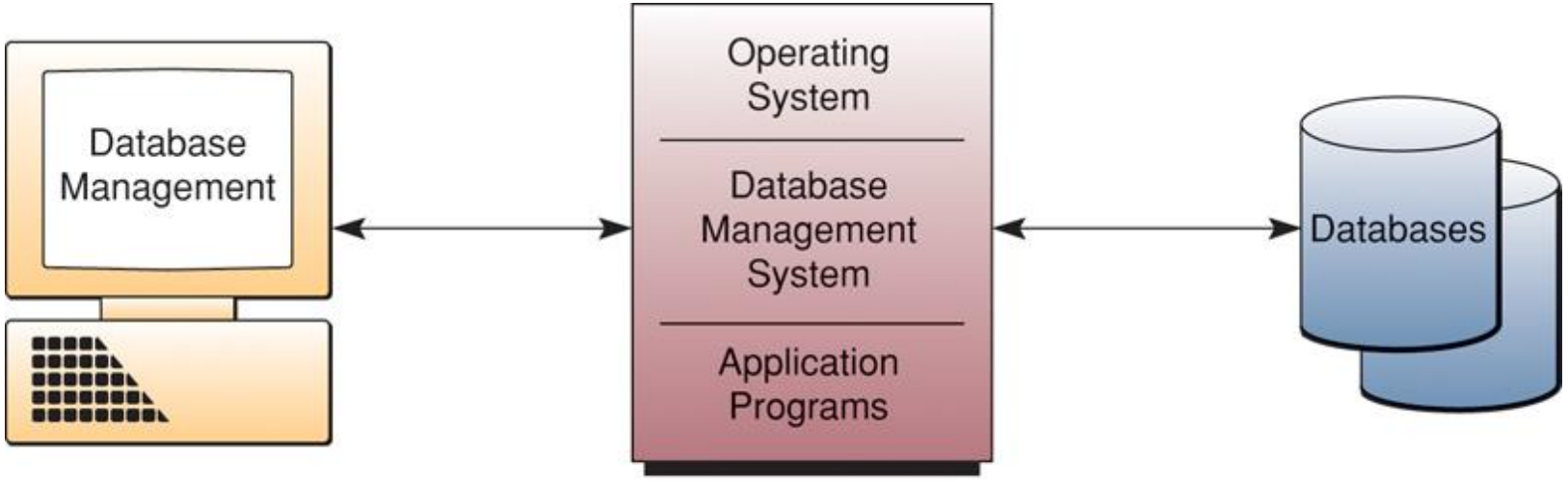
Copyright © 2019 by McGraw Hill Education. All rights reserved.

Common DBMS Software Components

Common DBMS Software Components	
• Database Definition	Language and graphical tools to define entities, relationships, integrity constraints, and authorization rights.
• Nonprocedural Access	Language and graphical tools to access data without complicated coding.
• Application Development	Graphical tools to develop menus, data entry forms, and reports.
• Procedural Language Interface	Language that combines nonprocedural access with full capabilities of a programming language.
• Transaction Processing	Control mechanisms to prevent interference from simultaneous users and recover lost data after a failure.
• Database Tuning	Tools to monitor and improve database performance.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

DBMS Major Functions



- Create: Database and Application Development
- Maintain: Database Maintenance
- Use: Database Interrogation

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Interrogation

SQL Queries

- Structured Query Language
- International standards
- In many DBMS packages
- Query form is

  SELECT…FROM…WHERE…

Operations Support Systems

A Sample Natural Language-to-SQL Translation for Microsoft Access

Natural Language

What Customers had no orders last month?

SQL

```
SELECT [Customers].[Company Name],[Customers].[Contact Name]
FROM [Customers]
WHERE not Exists {SELECT [Ship Name] FROM [Orders]
   WHERE Month {[Order Date]}=1 and Year {[Order Date]}=2004 and [Customers].
   [Customer ID]=[Orders].[Customer ID]}
```

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Interrogation

Boolean Logic

- Developed by George Boole
- Mid-1800s
- Used to refine searches
- Three logical operators: AND, OR, NOT

Example

- Cats OR felines AND NOT dogs OR Broadway

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Database Maintenance

Accomplished by

- Transaction processing systems
- Utilities and other applications, supported by DBMS
- Records new business transactions
- Updating and correcting data
  - Customer addresses

Copyright © 2019 by McGraw Hill Education. All rights reserved.

Application Development

DBMS tools

- 4GL programming language
- Built-in software development tools
- Data manipulation language (DML) statements
  - Eliminate conventional programming

Applications

- Data entry screens
- Forms
- Reports
- Web pages

Copyright © 2019 by McGraw Hill Education. All rights reserved.

RWC 3: Mergers Go More Smoothly

IT issues make or break mergers

Data Center feels impact

Knowledge lost if “acquired” IT people let go

Documentation necessary

Companies need to know what goes on in their data centers

Copyright © 2019 by McGraw Hill Education. All rights reserved.

RWC 4: Data Mining

Aggregate data for better decisions

Applebee's

- Uses back-of-house data to analyze food preparation
- Combined with front-of-house data
  - Determine time spent with customer
  - What is selling / what to order / what to promote

Travelocity

- 600,000 comments in e-mails and call notes
- Set up mining system to extract meaning

VistaPrint

- Analyzes customer online behavior

Copyright © 2019 by McGraw Hill Education. All rights reserved.

RWC 5: RedBus.In

Aggregator of Bus Travel Industry.

Automating the Bus Reservation System

The need was generated because of volatility of demand in the bus industry.

New system was to handle both reservation and bus-driver allocation.

Copyright © 2019 by McGraw Hill Education. All rights reserved.

RWC 6: Implementing HANA & BI

For full exploitation of data and analytics one requires three capabilities:

- Choosing the right data and manage multiple data sources.
- Need to build advanced models that turn the data into insights.
- Management must undertake a transformational-change program so that the insights translate into effective action.
- Case focuses on how USHA International implemented these three capabilities.

Copyright © 2019 by McGraw Hill Education. All rights reserved.