Discovering Computers Enhanced Edition ©2017

Tools, Apps, Devices, and the Impact of Technology

Chapter 12
Database

1. Databases, Data, and Information

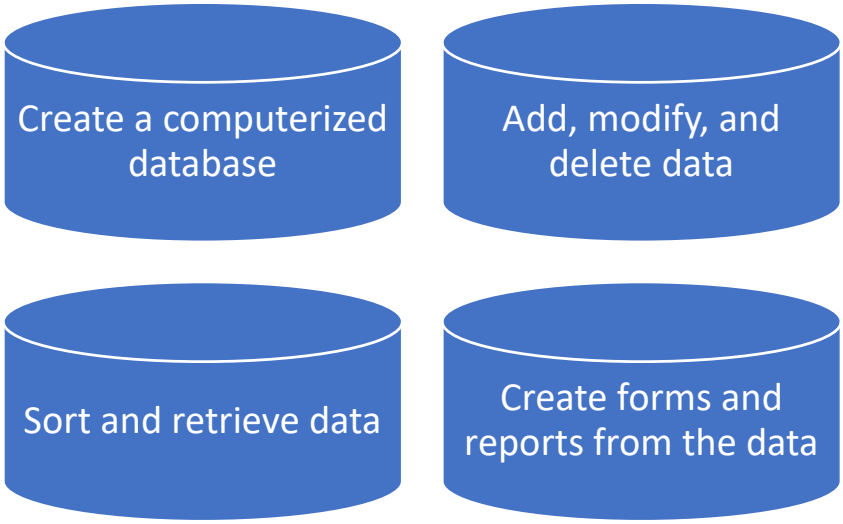
Database	Data	Information

• Collection of data organized in a manner that allows access, retrieval, and use of that data	• Collection of unprocessed items
	  • Text
	  • Numbers
	  • Images
	  • Audio
	  • Video
		• Processed data
		  • Organized
		  • Meaningful
		  • Useful

Computers process data in a database to generate information for users.

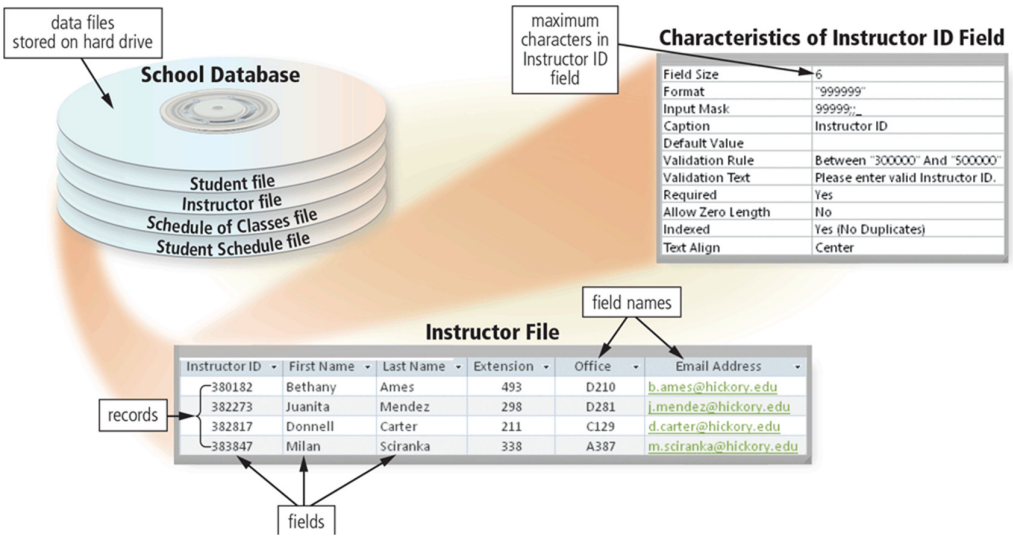
1. Databases, Data, and Information

- **Database software**, often called a **database management system** (**DBMS**), allows users to:



2. The Hierarchy of Data

- Data is organized in levels. Each higher level of data consists of one or more items from the lower level.
- A **database** contains a group of related **data files**. A data file contains **records**, a record contains **fields**, and a field is composed of one or more **characters**.



2. The Hierarchy of Data

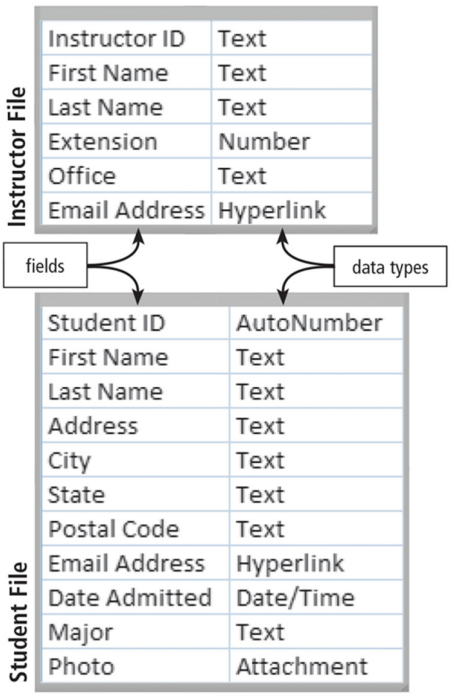
Characters, a **bit** is the smallest unit of data the computer can process. Eight bits grouped together in a unit constitute a **byte**. In the ASCII coding scheme, each byte represents a single **character**, which can be a number (4), letter (R), blank space (spacebar), punctuation mark (?), or other symbol (&).

Fields A field is a combination of one or more related characters or bytes and is the smallest unit of data a user accesses. A **field name** uniquely identifies each field.

A database uses a variety of characteristics, such as **field size** and **data type**, to define each field.
The **field size** defines the maximum number of characters a field can contain

The **data type** specifies the kind of data a field can contain and how the field is used.

2. The Hierarchy of Data



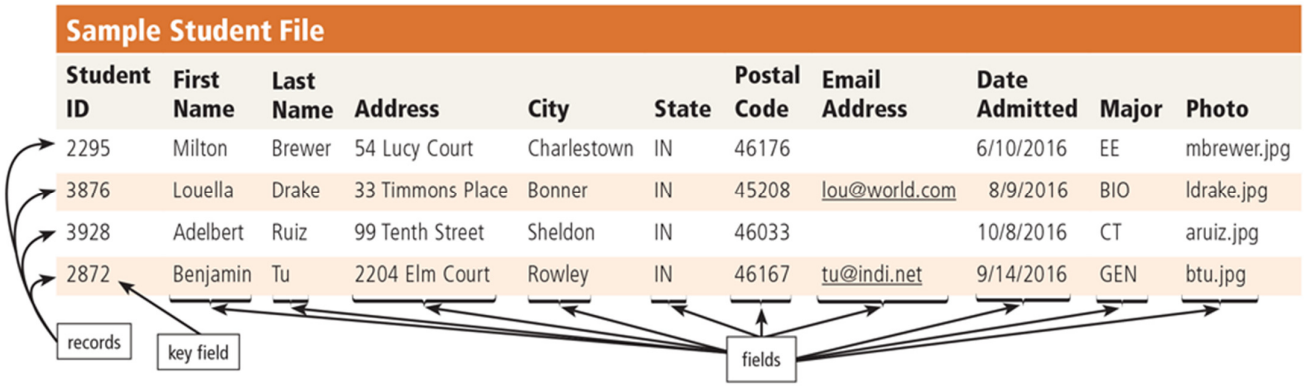
Common data types include the following:
- **Text**: Letters, numeric characters, or special characters
- **Number** (also called numeric values): Positive or negative numbers, and the number zero, with or without decimal points
- **AutoNumber:** Unique number automatically assigned by the DBMS to each added record, which provides a value that identifies the record (such as a student ID)
- **Currency:** Dollar and cent amounts or numbers containing decimal values
- **Date (**also called date/time): Month, day, year, and sometimes time
- **Memo** (also called long text): Lengthy text entries, which may or may not include separate paragraphs
- **Hyperlink:** Email address or web address that links to a webpage on the Internet or document on a network

2. The Hierarchy of Data

Records A record is a group of related fields.

A primary key is a field that uniquely identifies each record in a file. The data in a primary key is unique to a specific record.

Data Files A data file, often simply called a file, is a collection of related records stored on a storage medium, such as a hard drive, or on cloud storage.



Sample Student File

Student ID	First Name	Last Name	Address	City	State	Postal Code	Email Address	Date Admitted	Major	Photo
2295	Milton	Brewer	54 Lucy Court	Charlestown	IN	46176		6/10/2016	EE	mbrewer.jpg
3876	Louella	Drake	33 Timmons Place	Bonner	IN	45208	lou@world.com	8/9/2016	BIO	ldrake.jpg
3928	Adelbert	Ruiz	99 Tenth Street	Sheldon	IN	46033		10/8/2016	CT	aruiz.jpg
2872	Benjamin	Tu	2204 Elm Court	Rowley	IN	46167	tu@indi.net	9/14/2016	GEN	btu.jpg

3. File Maintenance

File maintenance refers to the procedures that keep data current.
File maintenance includes **adding records to**, **modifying records in**, and **deleting records from a file**.

Users add new records to a file when they obtain additional data that should be stored, such as data about a new student admitted to a school.

Generally, users modify a record in a file for two reasons:

(1) to correct inaccurate data or

(2) to update old data with new data, such as replacing a student's address when she moves to a new address.

When a record no longer is needed, a user deletes it from a file.

DBMSs use a variety of techniques to manage deleted or obsolete records. Sometimes, the DBMS removes the record from the file immediately, which means the deleted record cannot be restored. Other times, the record is **flagged, or marked**, so that the DBMS will not process it again.

4. Validating Data

Validation is the process of comparing data with a set of rules or values to determine if the data meets certain criteria. Many programs perform a validity check that analyzes data, either as you enter the data or after you enter it, to help ensure that it is valid.

Alphabetic/Numeric Check An alphabetic check ensures that users enter only alphabetic data into a field.

Range Check A range check determines whether a number is within a specified range.

Consistency Check A consistency check tests the data in two or more associated fields to ensure that the relationship is logical and their data is in the correct format.

Completeness Check A completeness check verifies that a required field contains data. For example, some fields cannot be left blank; others require a minimum number of characters.

Check Digit A check digit is a number(s) or character(s) that is appended to or inserted in a primary key value.

4. Validating Data

Table 11-1 Sample Valid and Invalid Data

Validity Check	Field(s) Being Checked	Valid Data	Invalid Data
Alphabetic Check	First Name	Karen	Ka24n
Numeric Check	Current Enrollment	24	s8q
Range Check	Per Credit Hour Fee	$220.25	$2,120.00
Consistency Check	Date Admitted, Birth Date	9/19/2016 8/27/1998	9/19/2016 8/27/2017
Completeness Check	Last Name	Gupta	
Other Check	Email Address	eg@earth.net	egearth.net

5. File Processing Systems and Databases

File Processing Systems

In the past, many organizations exclusively used file processing systems to store and manage data.

In a typical file processing system, each department or area within an organization has its own set of files. The records in one file may not relate to the records in any other file. Many of these systems have two major weaknesses: redundant data and isolated data.

ซ้ำซ้อน

- **Redundant data**: Because each department or area in an organization has its own files in a file processing system, the same fields are stored in multiple files. Duplicating data in this manner can increase the chance of errors.

โดดเดี่ยว

- **Isolated data:** It often is difficult to access data that is stored in separate files in different departments.

5. File Processing Systems and Databases

The Database Approach

When an organization uses a database approach, many programs and users share the data in the database.

Advantages of a Database Approach

- Reduced data redundancy
- Improved data integrity
- Shared data

ความสมบูรณ์

- Easier access
- Reduced development time

Disadvantages of a Database Approach

- Can be more complex than a file processing system
- Require more memory and processing power
- Data can be more vulnerable

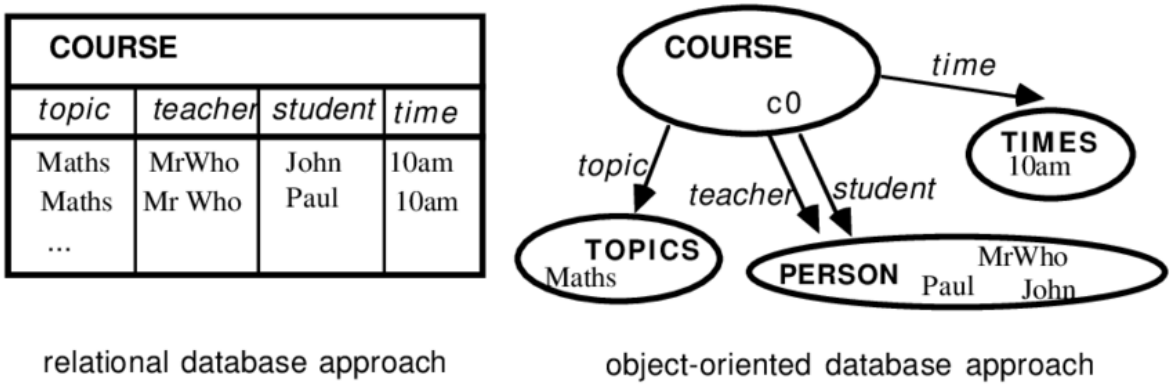
เสี่ยง, เป็นภัย

6. Types of Databases

A **data model** defines **how users view the organization of the data**. It does not define how the operating system actually arranges the data on the storage media. A database typically is based on one data model. Three popular data models in use today are relational, object-oriented, and multidimensional.

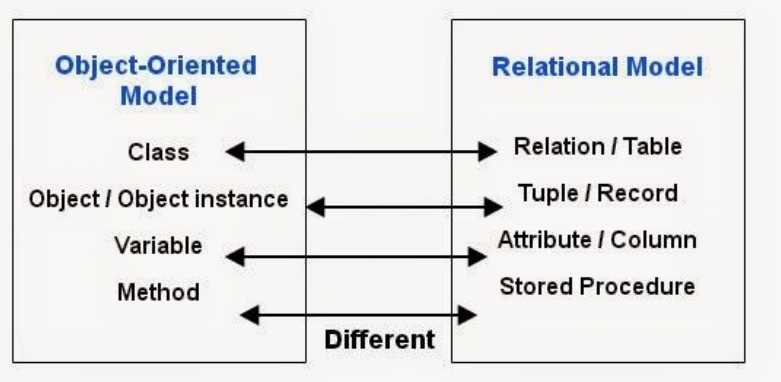
Relational Database A relational database is a database that stores data in tables that consist of rows and columns. Many organizations use relational databases for payroll, accounts receivable, accounts payable, general ledger, inventory, order entry, invoicing, and other business-related functions.

Object-Oriented Database An object-oriented database (OODB) stores data in objects. Examples of applications appropriate for an object-oriented database include media databases that store images, audio clips, and/or video clips.



relational database approach

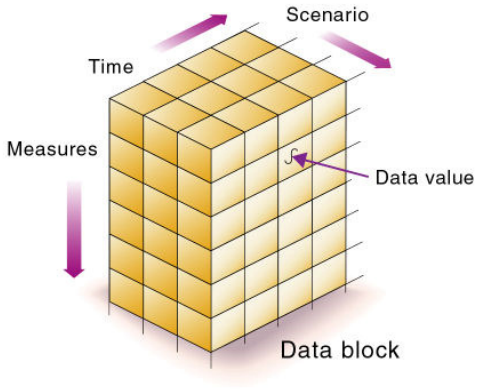
object-oriented database approach

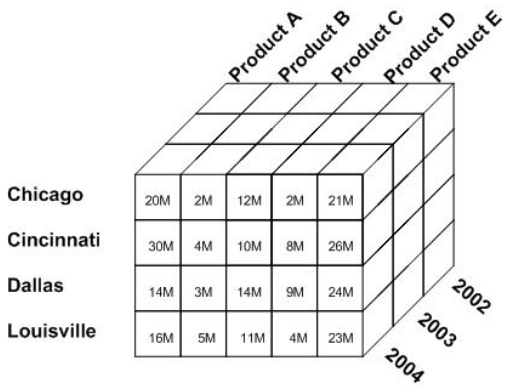


6. Types of Databases

Multidimensional Database A multidimensional database stores data in dimensions. Whereas a relational database is a two-dimensional table, a multidimensional database can store more than two dimensions of data. These multiple dimensions allow users to access and analyze any view of the database data.

One application that uses multidimensional databases is a **data warehouse**. A data warehouse is a huge database that stores and manages the data required to analyze historical and current transactions.



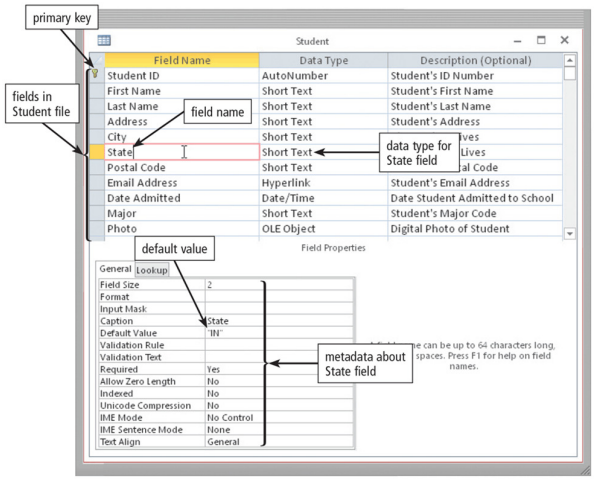


7. Database Management Systems

A database management system (DBMS), or database program, is software that allows you to create, access, and manage a database. Managing a company's databases requires a great deal of coordination.

The database administrator (DBA) is the person in the organization who is responsible for managing and coordinating all database activities, including development, maintenance, and permissions.

Data Dictionary A data dictionary, sometimes called a repository, contains data about each file in the database and each field in those files



7. Database Management Systems

File Retrieval and Maintenance A DBMS provides several tools that allow users and programs to retrieve and maintain data in the database. To retrieve or select data in a database, you **query** it. A **query** is a request for specific data from the database.

The four more commonly used are
query languages,
query by example,
forms,
and report writers.

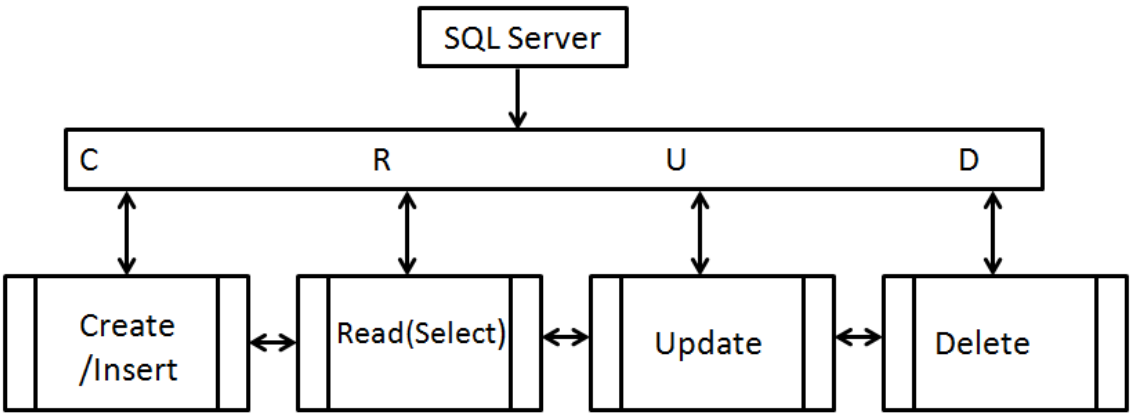
7. Database Management Systems

- A **query** is a request for specific data from the database
- A **query language** consists of simple, English-like statements that allow users to specify the data to display, print, store, update, or delete
- **Structured Query Language** (**SQL**) is a popular query language that allows users to manage, update, and retrieve data

```
SELECT CLASS_TITLE, CLASS_SECTION,
 MAXIMUM_ENROLLMENT - CURRENT_ENROLLMENT AS SEATS_REMAINING
FROM SCHEDULE_OF_CLASSES, CLASS_CATALOG
WHERE SCHEDULE_OF_CLASSES.CLASS_CODE = CLASS_CATALOG.CLASS_CODE
ORDER BY CLASS_TITLE
```

Class Title	Class Section	Seats Remaining
Algebra 1	51	14
Art Appreciation	52	19
English Composition 1	02	5
Introduction to Sociology	01	14

Structured Query Language (SQL)



- `SELECT ... FROM ... WHERE ...`
- `INSERT INTO ... VALUES ...`
- `UPDATE ... SET ... WHERE ...`
- `DELETE FROM ... WHERE ...`

Results | Messages

	ID	Fruit_Name	Fruit_Color
1	1	Banana	Yellow
2	2	Apple	Red
3	3	Lemon	Yellow
4	4	Strawberry	Red
5	5	Watermelon	Green
6	6	Lime	Green

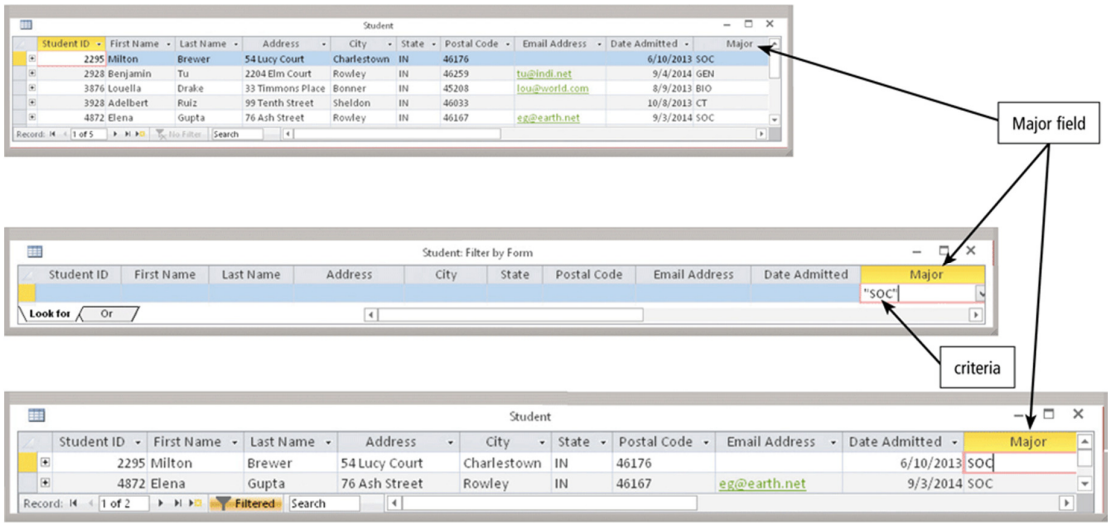
```
SELECT * FROM Fruits WHERE Fruit_Color='Red'
```

Results | Messages

	ID	Fruit_Name	Fruit_Color
1	2	Apple	Red
2	4	Strawberry	Red

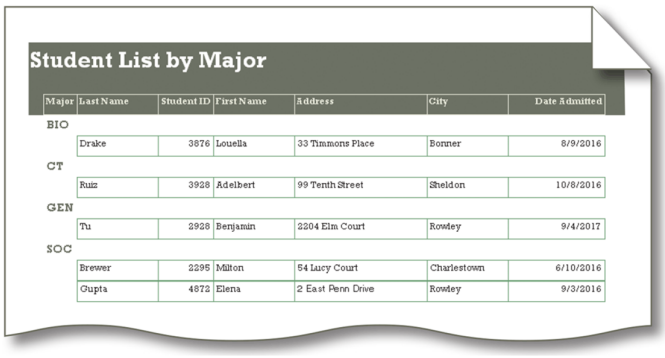
7. Database Management Systems

- Most DBMSs include **query by example** (**QBE**), a feature that has a graphical user interface to assist users with retrieving data



7. Database Management Systems

- A **form** is a window on the screen that provides areas for entering or modifying data in a database
- A **report writer** allows users to design a report on the screen, retrieve data into the report design, and then display or print the report



Student List by Major

Major	Last Name	Student ID	First Name	Address	City	Date Admitted
BIO						
	Drake	3876	Louella	33 Timmons Place	Bonner	8/9/2016
CT						
	Ruiz	3928	Adelbert	99 Tenth Street	Sheldon	10/8/2016
GEN						
	Tu	2928	Benjamin	2204 Elm Court	Rowley	9/4/2017
SOC						
	Brewer	2295	Milton	54 Lucy Court	Charlestown	6/10/2016
	Gupta	4872	Elena	2 East Penn Drive	Rowley	9/3/2016

7. Database Management Systems

Data Security

A DBMS provides means to ensure that only authorized users access data

อนุญาต

สิทธิ

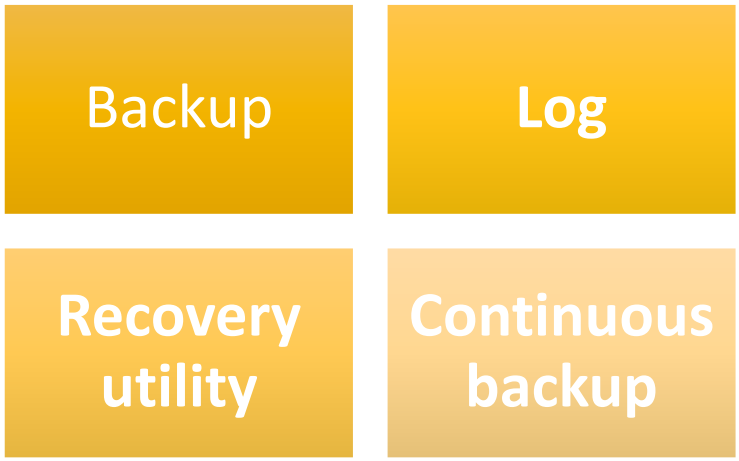
- Access privileges
- Principle of least privilege policy

การอนุญาตให้ผู้ใช้เข้าระบบหรือเข้าถึงข้อมูลแตกต่างกันตามภาระหน้าที่ความรับผิดชอบ

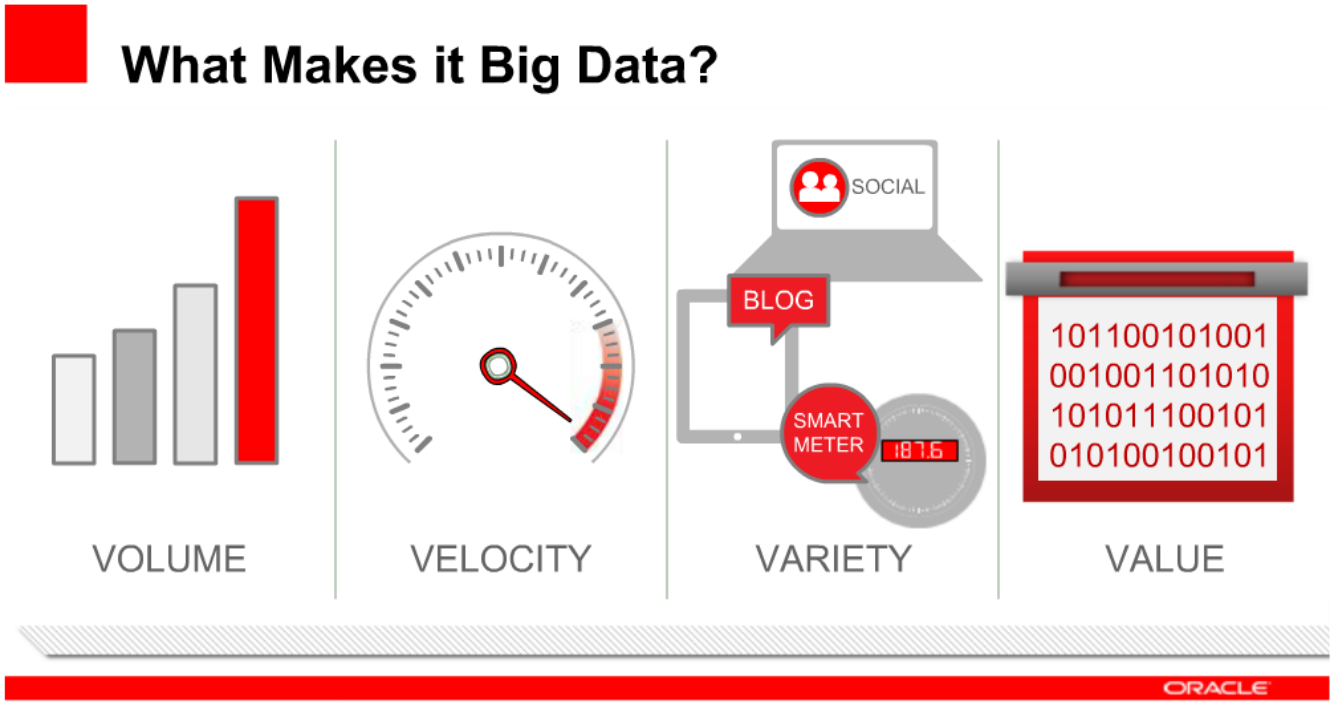
7. Database Management Systems

Backup and Recovery

- A DMBS provides a variety of techniques to restore the database to a usable form in case it is damaged or destroyed



Big Data



Goal of Big data : Find useful pattern or models in data

Big Data ที่วิเคราะห์จากการเลือกดูหนังที่ผ่านมาว่า ผู้ชมน่าจะอยากดูเรื่องใดเป็นเรื่องต่อไป แล้วนำเสนอได้ตรงหรือใกล้เคียงกับความชอบและรสนิยมของผู้บริโภคมากขึ้นเรื่อย ๆ ส่งผลต่อประสบการณ์และความพึงพอใจ ทำให้มีผู้สมัครสมาชิกมากกว่า 137 ล้านราย (ข้อมูล ณ สิ้นปี 2018)