

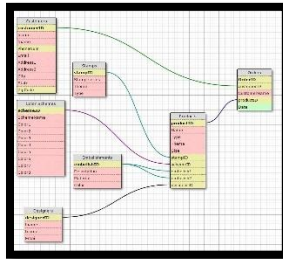


private information against the will of their citizens. In the U.S., some are calling on businesses to establish a Code of Ethical Practices for data management. Where do we draw the line between what is acceptable data collection for businesses and governments and what is not?

Another ethical dilemma related to databases has to do with the risk of private data falling into the hands of hackers and criminals. Information theft has become a multibillion-dollar industry with hackers utilizing tools to break through the security of databases in order to steal valuable information and individuals' identities. Information security is an important consideration when designing databases.

## Lesson 12.1: Databases and Big Data

### Lesson 12.1 Introduction



"CardsDB" by zedic.judd is licensed under CC BY 2.0

A **database** is a collection of data organized to meet users' needs. Databases have made it possible for geneticists to map the structure of DNA and for all scientists to share their research; without databases, the human genome project would not be possible. Databases are the unsung heroes of the Internet. Without databases, there would be no Google, Amazon, eBay, or Facebook. Databases and the systems that manage them are at the foundation of all successful information services, and they are becoming increasingly important as people generate more and more data for businesses and organizations to analyze.

Databases range in size and complexity from simple personal information management (PIM) software to large industrial databases that manage petabytes of data. The larger a database, the more time is required to define the relationships between the data and its underlying structure.

Years ago, corporate data, such as inventory and sales records, was accessed only by select individuals and divisions within an organization. Data was collected into a single ledger and copied and shared as needed. Today, organizations use centralized databases and large data centers accessed from many locations and systems over networks and the Internet. A grocery store chain, for example, runs multiple applications, including transaction processing, inventory ordering, and payroll, from one centralized database system. This centralized database approach allows the chain's management to obtain and analyze customer receipts from hundreds of stores.

The advantages of using a centralized database system are summarized in the following table. Although databases and data centers can be costly to set up and maintain, companies and organizations find that the advantages more than justify the costs.

#### Centralized Database Advantages

Advantage	Explanation
Reduced data redundancy	Data redundancy is the duplication of data across multiple locations.
Improved data integrity	Data integrity refers to the quality and accuracy of the data.
Easier modification and updating of data	Data is changed and updated in only one location.

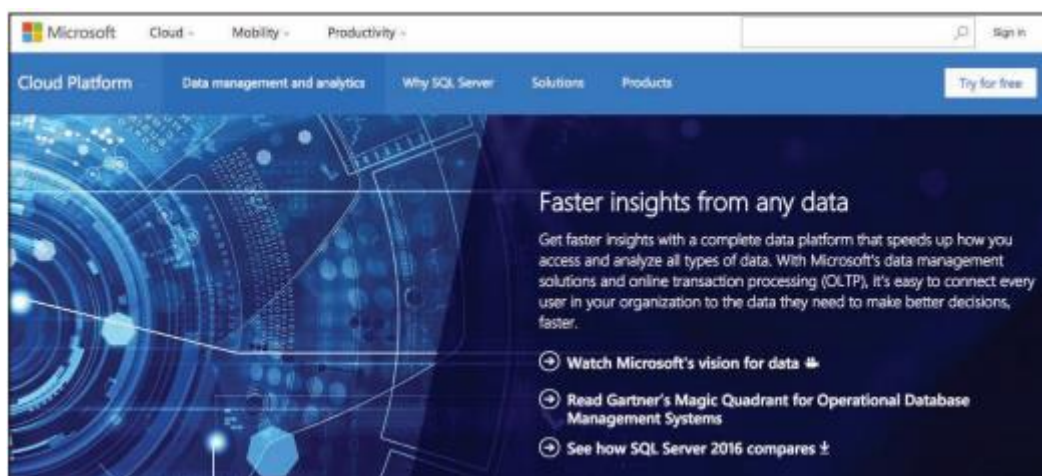
Advantage	Explanation
Easier access to reports and data; convenience of shared data access	Many of today's databases produce reports accessible on the web from any Internet-connected device.
Improved data security	It is easier to implement security measures in a centralized system.
Improved data storage efficiency	Having data in one location reduces storage needs.

### Reading: Database Management System (DBMS)

A **database management system or DBMS** consists of a group of programs that manipulate the data within a database, providing an interface between the database and the user or between the database and application programs.

#### Why This Matters

Individuals use personal databases to store addresses and contact information, track important appointments, inventory valuables for possible insurance claims, catalog music and book collections, and store and manage other lists of personal data. All of us interact with databases over the Internet. In fact, the web itself could be viewed as the world's largest database management system. Business systems depend on databases for daily operations and strategic planning. Database management systems serve as the user interface to data, providing data access and processing capabilities.



Courtesy of Microsoft Corporation. Fair Use

#### Essential Information

Database management systems range from small, inexpensive software packages to sophisticated systems costing hundreds of thousands of dollars. A few popular alternatives include flat file, single user, multiuser, and general-purpose and special-purpose systems. Open source databases are also available.

A flat file database stores database records in a plain text file. This method of data storage typically stores each record as a line of text and uses commas, tabs, or other indicators within the line to separate the fields within the record. The most common type of flat file format is called CSV for comma-separated values. Many specialized apps allow users to export their data into a CSV file for transfer into other similar apps. For example, Gmail users can export their contacts into a CSV file that can then be opened in Microsoft Excel or imported into another Contacts app like Microsoft Outlook.

Databases for personal computers are most often designed for a single user. Only one person can use the database at any time. Microsoft Outlook and Quicken are examples of popular special-purpose databases. These DBMSs are used to store and manipulate personal data. Outlook organizes contacts, calendars, email, and to-do lists, and Quicken is used for personal financial data.

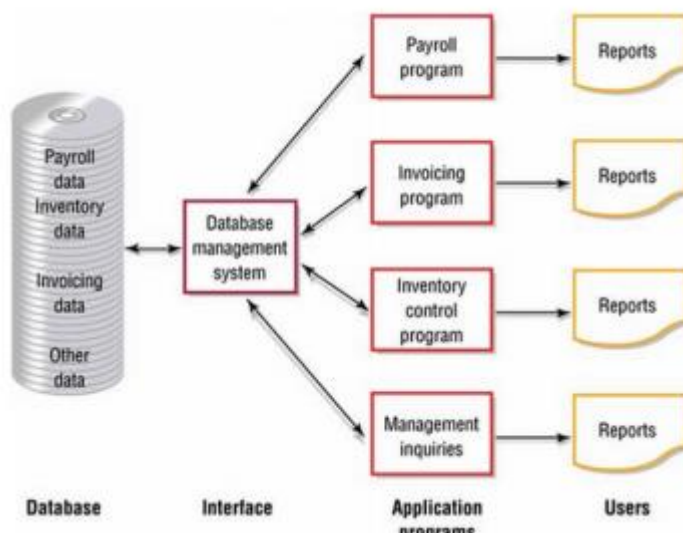


Google. Fair Use

Most small, medium, and large businesses require multiuser DBMSs that allow multiple employees to access and edit data simultaneously. These are typically referred to as database development platforms because they allow you to create a database and a DBMS for accessing and manipulating data in the database. Multiuser DBMSs utilize a unique manner of reading and writing data to provide users with simultaneous access to a shared database without the danger of overwriting each other's data. These databases often serve as the source of information for corporate applications like payroll, invoicing, and inventory (see figure below).

Microsoft Access is one of the smaller database development platforms available for use by individuals as well as small businesses and organizations. Large corporations that manage huge quantities of data invest significantly in industrial database systems. IBM DB2, Microsoft SQL Server, and Oracle are much more powerful and expensive, and they accommodate dozens, hundreds, and even thousands of simultaneous users at multiple locations. The Internet provides the network connection that allows users around the world to access one common database and that makes it possible for data to be collected from nearly anywhere. A distributed database takes available data stored in multiple locations and makes it appear as a single collection.

Oracle, Sybase, and IBM databases are examples of general-purpose database management systems that can be used for many different types of applications. Microsoft is currently the market leader in this category, with Oracle following in a close second. Like many technologies, database services are moving to the cloud. Internet companies like Amazon Web Services are becoming major players in the database space alongside traditional companies like Microsoft that are also developing cloud-based database services like Microsoft Azure.



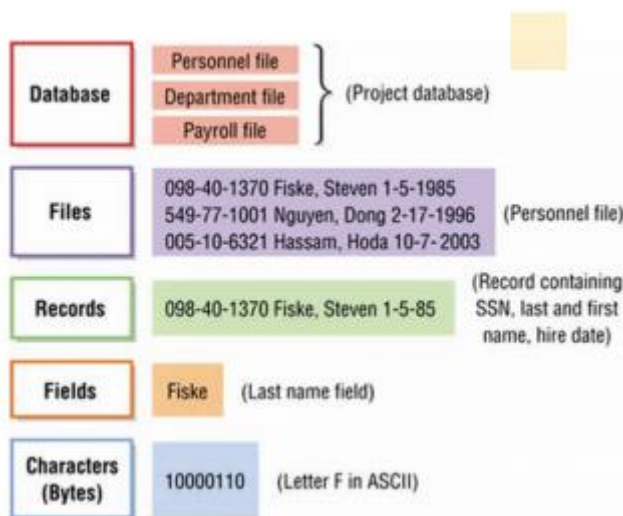
Authored by Cengage Learning. Creative Commons Attribution 4.0 International (CC BY 4.0)

### Reading: Data Hierarchy

**Data hierarchy** refers to the manner in which data in a database is organized into sequential levels of detail.

#### Why This Matters

Drilling down is a popular expression that refers to viewing the details related to some circumstance. For example, if you heard that the U.S. was threatened by a hurricane, you would want to drill down into that information to find out what region of the country would be impacted. You might drill down further to track the path and intensity of the storm. Database systems support these types of inquiries by arranging data in a hierarchy. A database hierarchy typically includes levels of data—beginning with characters, numbers, or other digital information—at the lowest level. These combine to make fields, which combine to make records, which combine to make tables, which combine to create the database, as shown in the following figure.



Authored by Cengage Learning. Creative Commons Attribution 4.0 International (CC BY 4.0)

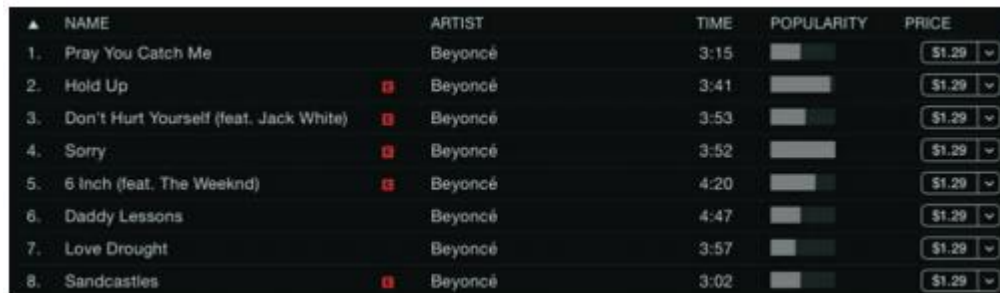
## Essential Information

Recall that in a computer, a byte is used to represent a character, which is the basic building block of textual information. Characters can be uppercase or lowercase letters (A, a, B, b, C, c,... Z, z), numeric digits (0, 1, 2,...9), or special symbols (![+][-]/€¥...).

In a database, characters are combined to form a field, the smallest practical unit in most databases. A field is typically a name, number, or combination of characters that in some way describes an aspect of an object (an individual, a song in a music library, an item in inventory, a photograph in an album) or activity (a business transaction, an interaction). Every field has a field name and can have either a fixed or a variable length. For example, songnumber might be the name of a field used to identify songs in a music library that requires an eight-digit code. On the other hand, artist\_name might be a field that can contain varying length content such as Beyoncé (7 characters), or P!nk (4 characters). Database syntax typically requires that field names not include any spaces, so spaces may be omitted or replaced with an underscore character.

A collection of related fields that describe some object or activity is a record. You can create a more complete description of an object or activity by combining fields that represent various characteristics of objects or activities into records. For instance, a database record for music in the iTunes Store combines fields for Name, Time, Artist, Price, Release Date, Track #, and others to fully describe the track.

A collection of related records is a table, also called a file in some databases. In the above example, all iTunes music might be stored in a table named Music. Similarly, all iTunes movies might be stored in a separate but related table named Movies.



	NAME	ARTIST	TIME	POPULARITY	PRICE
1.	Pray You Catch Me	Beyoncé	3:15	<div style="width: 100%;"></div>	\$1.29
2.	Hold Up	Beyoncé	3:41	<div style="width: 100%;"></div>	\$1.29
3.	Don't Hurt Yourself (feat. Jack White)	Beyoncé	3:53	<div style="width: 100%;"></div>	\$1.29
4.	Sorry	Beyoncé	3:52	<div style="width: 100%;"></div>	\$1.29
5.	6 Inch (feat. The Weeknd)	Beyoncé	4:20	<div style="width: 100%;"></div>	\$1.29
6.	Daddy Lessons	Beyoncé	4:47	<div style="width: 100%;"></div>	\$1.29
7.	Love Drought	Beyoncé	3:57	<div style="width: 100%;"></div>	\$1.29
8.	Sandcastles	Beyoncé	3:02	<div style="width: 100%;"></div>	\$1.29

Apple Inc. Fair Use

At the highest level of this hierarchy is the database, a collection of integrated and related files or tables. Apple may store music data in one table, movie data in another, TV show data in another, and podcast data in yet another. The combination of these tables makes up the complete iTunes database.

Databases use entities, attributes, and keys to store data and information. An entity is a generalized class of people, places, or things (objects) for which data is collected, stored, and maintained. Entities are represented as tables in a database. Records are instances of an entity. For example, “tracks” would be an entity in a database, and Beyoncé’s song “Sorry” an instance of the tracks entity. An attribute is a characteristic of an entity such as the length of a track. The specific value of an attribute, called a data item – such as 3:52, can be found in the fields of the record describing an entity.

### Reading: Database Key

A **database key** is a field in a table used to identify a record, such as EmployeeNumber.

email	first_name	last_name	date_enrolled
kbaldauf@emailservice.com	Ken	Baldauf	January 20
jlennon@bestemail.com	John	Lennon	February 15
pcline@eldermail.com	Patsy	Cline	March 20
jbaez@fricknmail.com	Joan	Baez	April 1

Primary key field

Attributes (fields)

Entities (records)

Authored by Cengage Learning. Creative Commons Attribution 4.0 International (CC BY 4.0)

## Why This Matters

One of the biggest challenges in database management is ensuring that records aren't duplicated. For example, you might sign up to get email updates for a hot new product that is soon to be released. The next time you visit the product website, you might forget that you already signed up and submit the form again. You expect that system to be smart enough to realize that you already signed up and not to email you twice. If that system utilizes database keys and has identified the email address field as a primary key, it won't be sending you duplicate emails.

## Essential Information

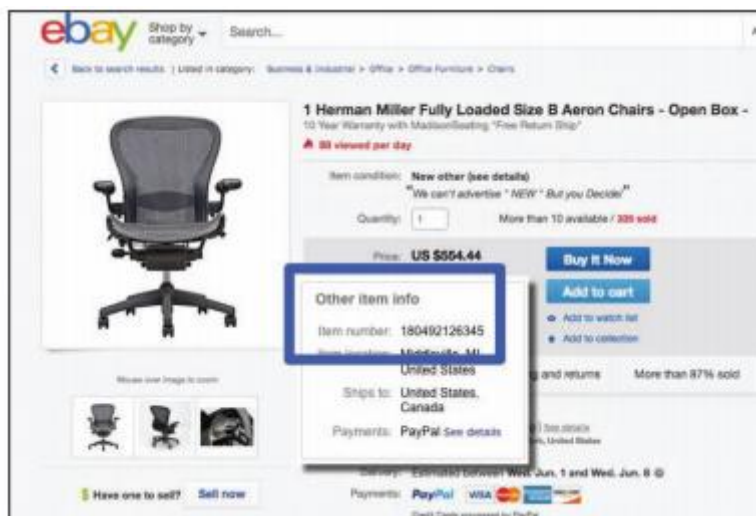
As you have learned, a record is made up of a collection of fields. A key is a field in a record that is used to identify the record, such as EmployeeNumber or Email\_address. A primary key is a field within a database table that uniquely identifies each record. Each record in a table must have a unique entry in the primary key field. The primary key distinguishes records so that they can be easily and accurately accessed, organized, and manipulated. For example, using employee numbers as the primary key ensures that every employee is identified only once, and using email addresses as the primary key for a mailing list ensures that no email address will be emailed more than once. The IRS uses Social Security numbers as the primary key to ensure that every person with a Social Security number is accounted for at tax time, and no one is counted twice.

In some cases, multiple fields may be combined as a primary key. For example, what would iTunes use as a primary key for the music files shown in this table?

### iTunes Music

Song	Artist	Album	Release date
Big Yellow Taxi	Joni Mitchell	Ladies of the Canyon	04/17/70
Big Yellow Taxi	Counting Crows/Vanessa Carlton	Hard Candy	07/09/02
Big Yellow Taxi	Amy Grant	House of Love	08/14/07

The Artist field might seem like a good option for a primary key, but each artist has many tracks. The Artist field is not unique for each record. The Song title is another option. But some songs are recorded by multiple artists, so the song title may not be unique to each database record. One solution is to use the combination of artist and song title to create the primary key. This way, Joni Mitchell's "Big Yellow Taxi" can be uniquely identified from other artists' recordings of "Big Yellow Taxi." If an artist has recorded the same song on multiple albums, it would be necessary to combine track title, artist, and album title as a primary key. Database designers often get around these types of challenges by assigning each record a unique ID number that acts as a primary key. For example, eBay identifies each item for sale with a unique item number.



eBay. Fair Use

### Reading: Relational Database

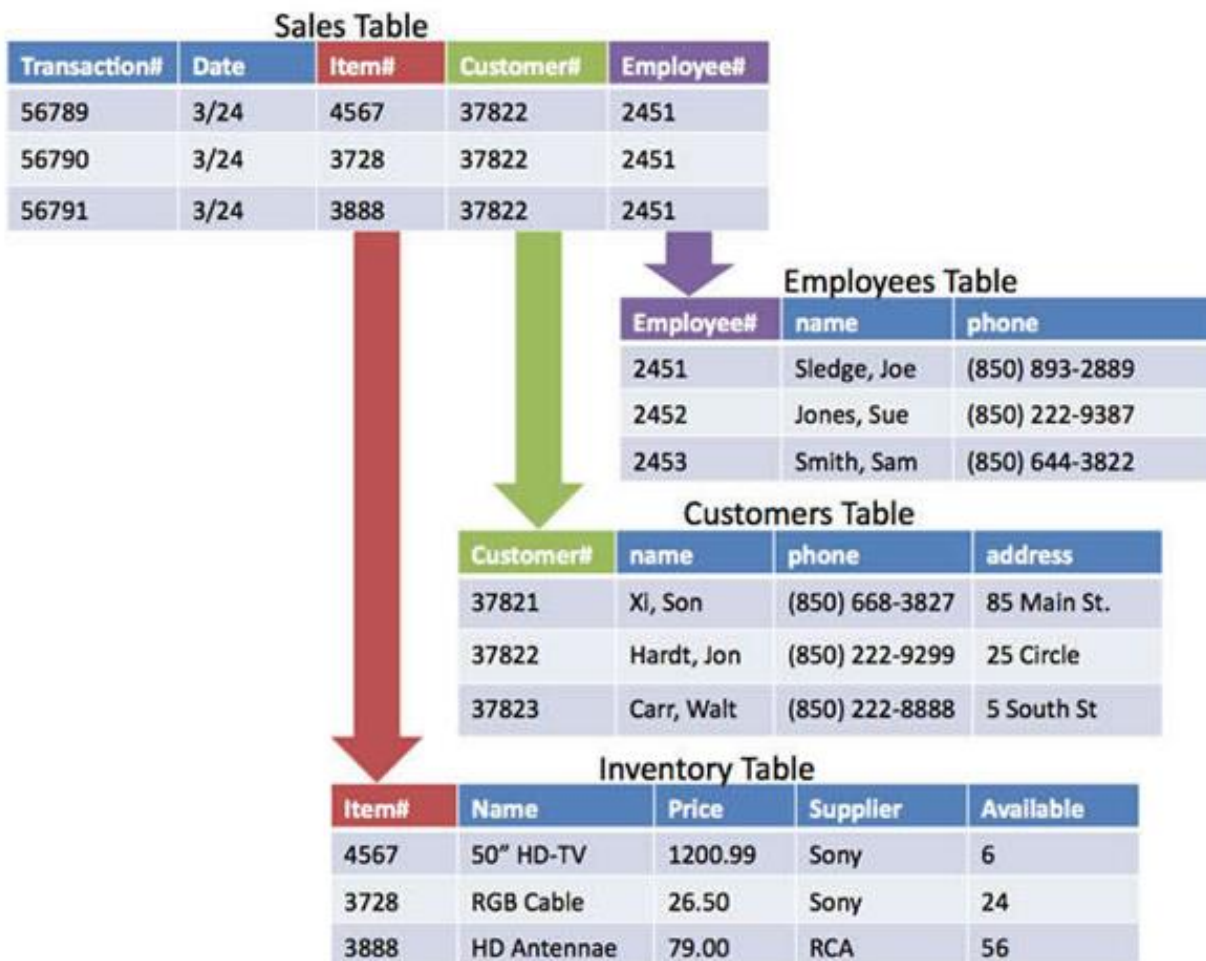
A **relational database** organizes data into multiple tables that are related by common fields.

#### Why This Matters

The data to be stored in a database often includes many categories of related information. For example, businesses typically store information on employees, inventory, customers, and sales. Storing all of this data in one large table would make for a huge and unwieldy database with many redundancies. Relational databases resolve this issue by storing data in separate tables according to logical categories, and relating the information by common fields.

Using the relationships among these tables, sales records can remain small and manageable.





Authored by Cengage Learning. Creative Commons Attribution 4.0 International (CC BY 4.0)

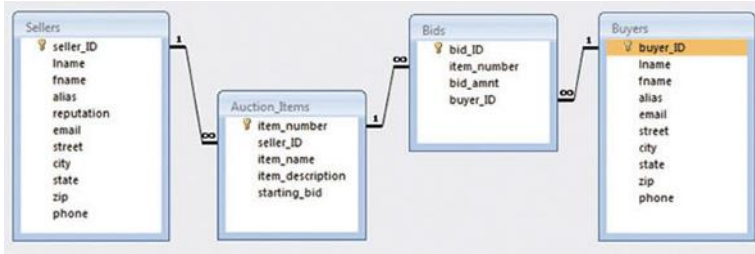
## Essential Information

The structure of the relationships in most databases follows a logical model. Many types of database models have been used over the years, including hierarchical, network, relational, object-oriented, and object-relational. The object-oriented model encapsulates data and database functionality together in software objects. Object-relational databases combine features of the object-oriented model with the relational model. The relational model is by far the most popular.

The overall purpose of the relational model is to describe data using a standard tabular format. In a relational database, all data elements are placed in two-dimensional tables organized in rows and columns, simplifying data access and manipulation. Notice in the figure above that the tables are related by common fields: Item#, Customer#, and Employee#. Having tables related by common elements allows them to be linked to produce useful information. This example is simplified to make it easy to understand. Typical business databases include dozens of tables with hundreds of relationships.

Relationships between tables can be illustrated through an entity relationship diagram. View the entity relationship diagram for an online auction database provided below. Notice that there are four tables in this database: Sellers, Auction Items, Bids, and Buyers. The fields included in each table are listed in the boxes. The key icon indicates the primary key for each table.

An online auction website might depend on four database tables, all related by common fields, as shown in this entity relationship diagram.



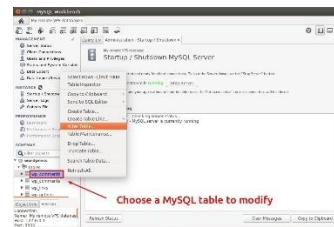
Courtesy of Microsoft Corporation. Fair Use

Notice the lines connecting the tables in the figure. These represent the relationships between the tables and indicate which fields in the tables hold common information. For example, the seller\_ID field is used to connect the Sellers table to the Auction\_Items table.

Also notice the 1 and infinity symbols used on the relationship lines. These indicate a special kind of relationship called a one-to-many relationship. For example, the relationship between the Sellers table and the Auction\_Items table has a 1 on the Sellers side. This is because there is only one record for every seller\_ID in the Sellers table, which is because seller\_ID is the primary key. However, there is an infinity sign over the Auction\_Items side of the relationship because each seller can have a limitless number of items up for auction—the seller\_ID is not a primary key in the Auction\_Items table. One-to-one and many-to-many relationships can also exist, depending on whether the field used is a primary key or not.

## Lesson 12.2: Database Tools

### Lesson 12.2 Introduction



"MySQL Workbench - Modify MySQL Table" by xmodulo is licensed under CC BY 2.0

**Database tools** include software and techniques for analyzing, maintaining, and manipulating data in a database.

The power of a database lies in the tools utilized to manage, maintain, and manipulate the data. These tools enable investors to earn millions of dollars by buying and selling stocks at the most opportune moment. Or, they might assist you in finding great new music based on your general listening trends. The most fundamental database tools are schemas, data dictionaries, and data manipulation languages such as SQL.

Database tools exist to ensure the collection of high-quality data and to manipulate the data for some purpose. High-quality data is nonredundant, flexible, simple, and adaptable to a number of different applications. Data analysis is the process of evaluating data to identify problems with the content of a database.

The process of correcting data problems or anomalies is called normalization. It ensures that the database contains good data. Normalization usually involves breaking one table into two or more tables in order to correct a data problem or anomaly. Normalization is a very important technique in database management and is yet another advantage to using a relational database.

Data integrity refers to the quality of data: the degree to which it is accurate and up to date. There are many possibilities for inaccuracies in today's information-rich society. Some people have been unable to get credit cards, car loans, or home mortgages because data stored about them by credit bureaus was inaccurate. In some cases, a retail store may report that someone didn't pay his or her bills, when the bills were paid in full. These types of errors are caused by inaccurate data being entered into a database. The results are inaccurate output. This is sometimes referred to as garbage in/garbage out (GIGO). Computer systems are often blamed for such errors, but in reality, the errors are more often human.

Not only are tools designed to analyze and normalize data, but tools are also used to query, filter, and organize data to meet particular needs. Data manipulation languages, the most popular of which is SQL, are the main tools of those who work with databases.

### **Reading: Schema**

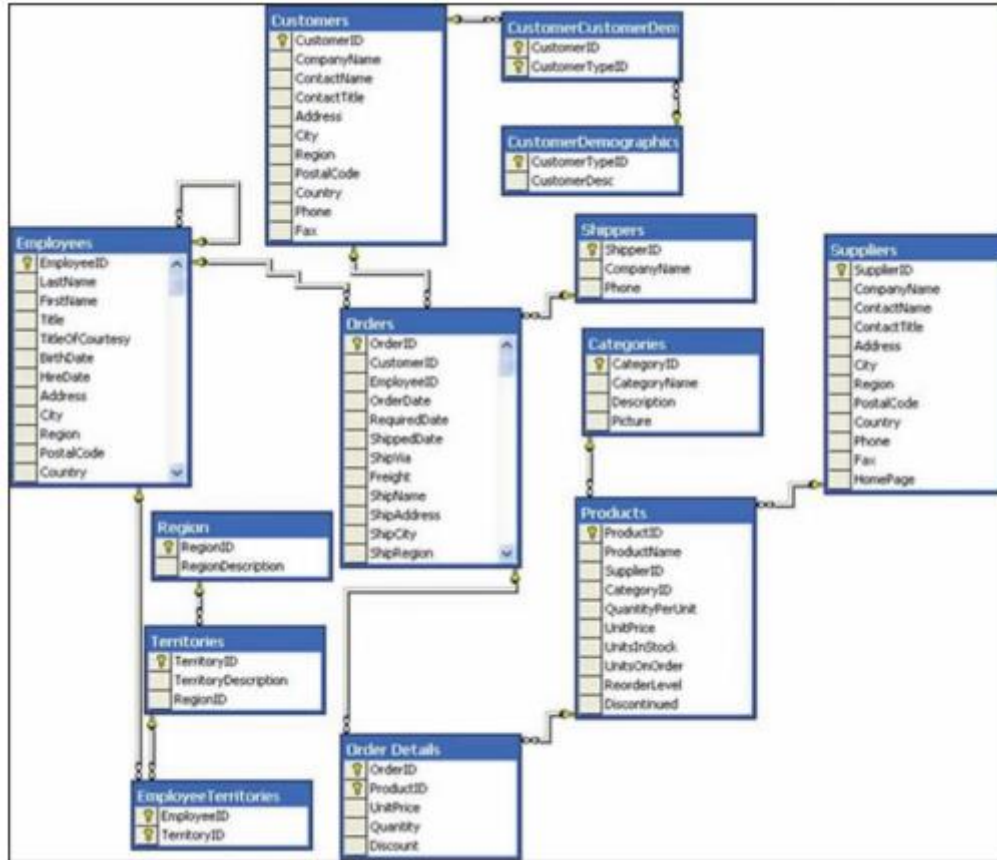
A **database schema** is a graphical representation of the structure of a database.

#### **Why This Matters**

The organization of data in a database can quickly become very complicated. Schemas allow database administrators and users to visualize the structure of a database and understand the relationships between data and tables.

#### **Essential Information**

One of the first steps in creating a database is to outline the logical and physical structure of the data as well as the relationships among the data in the database. This description is called a schema (as in schematic diagram). A schema can be part of the database or a separate schema file. The DBMS can reference a schema to find where to access the requested data in relation to another piece of data. Schemas are entered into the DBMS (usually by database personnel) via a data definition language. A data definition language (DDL) is a collection of instructions and commands used to define and describe data and data relationships in a specific database.



Courtesy of Microsoft Corporation. Fair Use

**Reading: Data Dictionary**

A **data dictionary** provides a detailed description of each field and table in a database.

**Why This Matters**

Data in a database is typically accessed by different individuals and software. A data dictionary provides specific details on the requirements of each field and table in a database, how they may be accessed, and the specific parameters and requirements imposed on the data they hold.

NORTHWESTERN MANUFACTURING	
PREPARED BY:	D. BORDWELL
DATE:	04 AUGUST
APPROVED BY:	J. EDWARDS
DATE:	13 OCTOBER
VERSION:	3.1
PAGE:	1 OF 1
DATE:	13 OCTOBER
DATA ELEMENT NAME:	PARTNO
DESCRIPTION:	INVENTORY PART NUMBER
OTHER NAMES:	PTNO
VALUE RANGE:	100 TO 5000
DATA TYPE:	NUMERIC
POSITONS:	4 POSITIONS OF COLUMNS

Authored by Cengage Learning. Creative Commons Attribution 4.0 International (CC BY 4.0)

## Essential Information

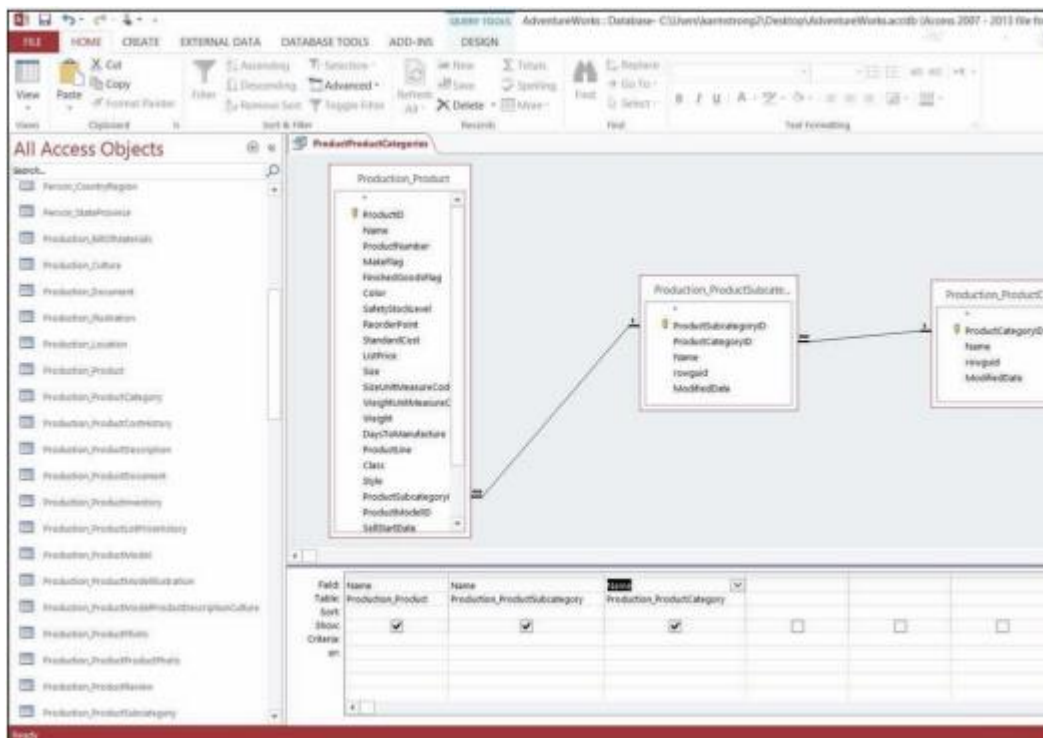
An important step in creating a database is to establish a data dictionary, a detailed description of all data used in the database. The data dictionary includes information such as the name of the data item, who prepared the data, who approved the data, the date, a description, other names that may be used to refer to the data, the range of values for the data, the data type (numeric or alphanumeric), and the number of positions or amount of space needed for the data. A data dictionary provides standard definitions of terms and data elements that can be referenced by programmers, database administrators, and users to maintain data integrity.

## Reading: Structured Query

**Structured Query Language (SQL)** is a popular data manipulation language used by the vast majority of database programmers and administrators for manipulating data to meet the needs of the users.

## Why This Matters

The power of a database lies in the ability to sift, sort, and query data to answer questions and discover patterns. For example, it is useful to discover which are the most profitable stores in your retail chain, who are the least productive employees, what airline offers the best deal on a particular route, which song is most popular, and what direction the economy is headed. Queries are designed and query results are generated through the use of data manipulation languages such as SQL.



Courtesy of Microsoft Corporation. Fair Use

## Essential Information

Once a DBMS has been installed and a database and table(s) created, the data can be accessed and manipulated using specific commands in various programming languages and a data manipulation language. A data manipulation language (DML) is a specific language provided with the DBMS that

allows database users to access, modify, and make queries about data contained in the database and to generate reports.

In 1986, the American National Standards Institute (ANSI) adopted Structured Query Language (SQL, also pronounced “sequel”) as the standard query language for relational databases. Today, SQL is an integral part of popular databases on both mainframe and personal computers. If you visit the Books section of Amazon.com and enter a search for J. K. Rowling, when you click the search button, it is likely that software running on Amazon’s servers issues an SQL statement such as

```
SELECT * FROM BOOKS WHERE AUTHOR = 'J. K. Rowling'
```

which would populate the webpage with all records from the BOOKS table that have J.K. Rowling listed in the author field.

SQL is actually a programming language, but it is easy for nonprogrammers to understand and use, thanks to its English-like commands. Programmers can use SQL on systems ranging from PCs to the largest servers. SQL statements also can be embedded in many programming languages, such as the widely used Javascript and PHP languages.

### Examples of SQL Commands

SQL Command	Description
SELECT ClientName, Debt FROM Client WHERE Debt>1000	This query displays all clients (ClientName) and the amount they owe the company (Debt) from a database table called Client, for clients that owe the company more than \$1000 (WHERE Debt>1000).
SELECT ClientName, ClientNum, OrderNum FROM Client, Order WHERE Client.ClientNum = Order.ClientNum	This command is an example of a join command that combines data from two tables: the client table and the order table (FROM Client, Order). The command creates a new table with the client name, client number, and order number (SELECT ClientName, ClientNum, OrderNum).
GRANT INSERT ON Client to BGuthrie	This command is an example of a security command. It allows Bob Guthrie to insert new values or rows into the Client table.

Some commercial databases provide tools and wizards for automating the process of building SQL statements. Microsoft Access provides a Query tool and wizard where tables and fields are selected and criteria specified (see figure at top of page) to quickly generate SQL queries. There is also the ability to query by example (QBE), where a database entity is selected as an example of the criteria to be queried. Query wizards and QBE automate the process of writing SQL commands, making manipulating databases easier and faster for novices.

## Lesson 12.3: Business Databases

### Lesson 12.3 Introduction

Enterprise databases are large databases that function as the cornerstone of information systems in businesses and enterprises.

When it comes to big data, you need big technologies to manage it. Large businesses, referred to as enterprises, invest heavily in storing and processing the data that feeds numerous information systems. Data analytics support decision-making processes throughout an organization. A number of companies have arisen to assist businesses of all sizes in managing and analyzing large amounts of data.



Courtesy of National Energy Research Scientific Computing Center. Public Domain

Data is the lifeblood of the enterprise, and the flow of key information to the right people at the right time is essential to the success of a business. Key technologies are used to assist in accomplishing this goal. Big businesses utilize servers in data centers to store large quantities of business-related data. A data warehouse is used to collect all types of data from a variety of resources. Data mining is used to harvest specific information to support essential business decisions and strategies. A database administrator designs, implements, and maintains the database, and trains others in its use.

Amazon.com trusts its huge multi-terabyte database to support its online transactions. To meet Amazon's high demands, its database must provide personalized, rapid responses to customers. It is typical for huge databases to lag in response time, but not so with Amazon's. After greeting customers with a custom-designed webpage built on the fly, the database serves up product information, user reviews, lists of recommendations, and links to more products—all in the blink of an eye. The DBMS is optimized to provide both agility and power. When the customer is ready to make a purchase, the DBMS manages the entire process.

Businesses depend on many different types of database technologies to claim a competitive advantage. Of those available, the most powerful and popular are discussed in this section.

### Reading: Data Warehouse

A **data warehouse** is a very large database that holds important information from a variety of sources.

#### Why This Matters

In the corporate environment, data is generated by a variety of functional units within the business: accounting, marketing, sales, human resources, etc. Business executives combine data from many sources and analyze it to get the big picture of what is occurring in the business. Data warehouses merge data from many sources into one large database designed for analysis. Data warehouses are a crucial component of many business systems.

#### Essential Information

A data warehouse is a large database that holds information from a variety of sources. It is usually a subset of multiple databases maintained by an organization or individual. Data warehouses are used by many companies and organizations. A hardware store, for example, can use a data warehouse to analyze pricing trends. This can help the store determine what inventory to carry and what price to charge. With a data warehouse, all you have to do is ask where a certain product is selling well, and a colorful table showing sales performance by region, product type, and timeframe pops up on the screen. Similarly, some police departments use data warehouses to store records of events such as complaints, arrests, criminal summonses, shootings, homicides, and other incidents. The data warehouse is tapped to

provide real-time information for a centralized help desk that provides detectives with the data needed to solve crimes.

A data mart is a small data warehouse, often developed for a specific person or purpose. It can be generated from a data warehouse using a database management system.

### Reading: Data Mining

**Data mining** refers to the process of extracting information from a data warehouse or data mart.

#### Why This Matters

The amount of data collected by today's corporations is far too huge to be interpreted solely by humans. Computerized tools are needed to discover irregularities and trends in data. Data mining is the application of such tools.

#### Essential Information

Data mining is the process of extracting information from a data warehouse or a data mart. The DBMS can be used to generate a variety of reports that help people and organizations make decisions and achieve their goals. Data mining has been used in the airline-passenger profiling system to block suspected terrorists from flying. Data mining is also used by government agencies to detect patterns of terrorist activity.

In a business setting, data mining can yield outstanding results. Often called business intelligence or BI, a term first coined by a consultant at Gartner Group, the business use of data mining can help increase efficiency, reduce costs, or increase profits. The business-intelligence approach was first used by Procter & Gamble in 1985 to analyze data from checkout scanners. Today, most companies use data mining and business intelligence to interpret valuable information provided by collected data.

### Reading: Data Center



"cabinets" by bugeaters is licensed under CC BY 2.0

A **data center** is a climate-controlled room, building, or set of buildings that houses servers storing and delivering mission-critical information and services.

#### Why This Matters

The world economy is increasingly dependent on data. Much of the world's data is stored in data centers of various sizes. Data centers are getting media attention because they consume large amounts of energy. New green technologies are being applied to help decrease the size and energy demands of the data centers that store our rapidly increasing amounts of data. Most of the energy required for a data center goes to keeping servers cool. Facebook built its European data center in Lulea, Sweden, on the edge of the Arctic Circle where it benefits from cheap electricity and year-round cool temperatures.



**Essential Information**

Corporate databases and database management systems are typically housed in facilities called data centers. A data center is a climate-controlled room, building or set of buildings that house servers storing and delivering mission-critical information and services. Data centers of large organizations are often distributed among several locations. Microsoft constructed a 400,000-square-foot data center in San Antonio, at a cost of \$550 million. Google constructed a \$600-million mega-data center in Lenoir, North Carolina, and a \$750-million data center in Goose Creek, South Carolina. The Range International Information Hub in LangFang, China is currently the world's largest data center, covering 331 acres. Clearly, storing and managing data is a serious business.

The number of data centers being constructed for big companies like Google, Microsoft, Amazon, Apple, and Facebook is rapidly increasing as computer users are storing increasing amounts of data in the cloud. Cloud computing is gradually moving all data off individual computers and onto Internet-accessed servers maintained in data centers.

While corporate and personal data sits in large supercooled data centers, the people accessing that data are typically in offices spread across the country or around the world. In fact, in the near future, the only personnel on duty at data centers may be security guards. Data centers are approaching the point of automation, where they can run and manage themselves while being monitored remotely. This is referred to as a "lights out" environment.