Unit III: Using Technology Effectively & Responsibly

Chapter 13: Data and Big Data

13.1: Goal

In this chapter, you will explore how data is collected, stored, analyzed, and used to support decision-making. You will examine the growing role of big data in business, healthcare, education, and science while considering the opportunities, challenges, and ethical issues associated with collecting and analyzing large amounts of information.

13.2: Objectives

- Define data and explain why it is valuable.
- Differentiate between structured and unstructured data.
- Explain how information is created from raw data.
- Describe the characteristics of big data.
- Identify common sources of data collection.
- Explain how organizations analyze data to make decisions.
- Discuss the benefits and challenges of big data.
- Recognize the importance of data privacy, security, and ethical data use.

13.3: Think About It

Every day, your smartphone records your location, streaming services remember what you watch, online stores recommend products based on previous purchases, fitness trackers count your steps, and navigation apps monitor traffic conditions.

Have you ever wondered how all of this information is collected and used?

The answer is **data**. Modern organizations collect enormous amounts of information, and learning how to organize, analyze, and protect that information has become one of the most important skills in today's digital world.

13.4: Opening Scenario

When Harry woke up this morning, he unlocked his phone and checked the weather. Next, he checked his calendar, then sent a text message to a friend.

On the way to work, he bought coffee and paid by swiping his credit card. At work, he swiped his Employee ID, then logged into his work computer to check his email.

Each of these activities generated data.

Most people create thousands of data points every day, often without realizing it. Organizations analyze this information to improve services, personalize recommendations, detect fraud, manage traffic, and support countless other applications.

13.5: Why This Chapter Matter

Data is often called the **new currency of the digital economy** because organizations rely on information to make decisions, improve products, predict trends, and solve problems. Businesses analyze customer purchases, hospitals use patient data to improve healthcare, scientists study climate data, and governments use information to plan transportation and public services.

Understanding data helps you become a more informed consumer, a more responsible digital citizen, and a more effective employee in almost any profession. Whether you work in healthcare, education, business, manufacturing, criminal justice, or information technology, you will encounter data every day.

By the end of this chapter, you will understand how data becomes useful information, why big data is transforming nearly every industry, and why privacy, security, and ethics remain essential whenever data is collected and analyzed.

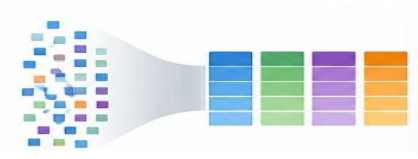
13.6: What Is Data?

Every time you send a text message, make an online purchase, watch a streaming video, or use a GPS navigation app, you create **data**. Computers, smartphones, sensors, and countless other digital devices generate enormous amounts of data every second.

Data consists of raw facts, observations, or measurements that have not yet been organized or interpreted. By itself, data may have little meaning. However, when data is organized, processed, and analyzed, it becomes **information** that people can use to make decisions. For example, imagine a weather station records the following temperatures during one week: 72, 74, 75, 71, 69, 73, 76

These numbers are **data**. After calculating the average temperature and identifying a warming trend, the results become **information** that helps meteorologists prepare weather forecasts. Organizations use information to answer questions, solve problems, identify patterns, and make better decisions.

Data vs. Information



Although people sometimes use the words *data* and *information* interchangeably, they have different meanings. **Data** consists of individual facts or measurements. **Information** is data that has been organized and interpreted so that it is useful: it has meaning and can support decision-making. **Figure**: Raw data is processed to create information.

For example, a grocery store records every purchase made during the day. Each transaction is data. After analyzing thousands of purchases, the store may discover that bottled water sales increase before major storms. That discovery is information that helps managers stock shelves more effectively.

Data vs. Information

DATA	INFORMATION
Raw facts	Organized and meaningful
Individual numbers or observations	Results of analyzing data
May have little meaning alone	Supports decisions
Example: Individual exam scores	Example: Class average and grade distribution
Example: Daily sales transactions	Example: Monthly sales report, video

Types of Data

Data appears in many different forms. Common types include numbers, text, images, video, dates and times.

A smartphone may collect GPS coordinates, photos, contact information, appointments, step counts, and messages.

13.7: How is Data Classified?

Data can be classified in several ways.

Qualitative or Quantitative

Qualitative data describes qualities or characteristics rather than numbers. Example: customer reviews, descriptions, synopsis of movie.

Quantitative data consists of numbers. Example: test scores, temperatures, steps for the day, account balance.

13.8: Structured and Unstructured Data

In the mail today, you got a credit card statement, a bank statement, a grade report from the college, and a letter from your aunt.

With the credit card statement, you quickly noted the amount due and the due date. You looked over the list of purchases to see if there was anything unexpected. You looked at the grade report and noted that the grades were the same ones you already checked online. It took longer to read the letter from your aunt.

Not all data is organized in the same way.

Structured data is organized into a predictable format, often using rows and columns within a database or spreadsheet. The credit card statement, bank statement, and grade report are **Structured data**. You knew exactly where to find the information you wanted.
Because structured data follows consistent rules, computers can search, sort, and analyze it efficiently.

Unstructured data does not follow a fixed format. The letter from your aunt is **Unstructured data**. You need to read the letter carefully to find the important information. Other examples include Emails, Social media posts, photographs, videos, PDF documents, and text messages.

Although unstructured data is more difficult to analyze, it represents the majority of digital information created today. Advances in artificial intelligence and machine learning have made it easier to identify patterns within large collections of unstructured data.

Structured vs. Unstructured Data

STRUCTURED DATA	UNSTRUCTURED DATA
Organized into tables	No fixed organization
Easy to search	More difficult to analyze
Stored in databases	Stored as documents, images, audio, or video
Student records	Emails
Inventory database	Social media posts
Payroll system	Digital photographs

13.9: Why Data is Important

- Organizations often describe data as one of their most valuable resources.
- Businesses analyze customer purchases to improve products.
- Hospitals study patient data to improve treatments.
- Scientists examine environmental data to better understand climate change.
- Schools use student performance data to improve instruction.
- Sports teams analyze player statistics to develop game strategies.
- Governments use census and transportation data to plan public services.

When accurate data is available, organizations can make better decisions, identify problems earlier, reduce costs, improve customer service, and discover new opportunities.

However, poor-quality data can lead to incorrect conclusions. For this reason, organizations work hard to ensure that their data is accurate, complete, timely, and secure.

13.10: Collecting and Managing Data

Where Does Data Come From?

Every interaction with technology has the potential to generate data. Whether you unlock your phone, make a purchase, scan a barcode, or stream a movie, computer systems record information that can be used to provide services, improve operations, or analyze trends. Organizations collect data from many different sources, including:

- Smartphones and tablets
- Websites and online stores
- Credit and debit card transactions
- Social media platforms
- Smart home devices
- Wearable technology
- Medical equipment
- Vehicles with built-in computers
- Environmental sensors • Business operations

Many of these systems collect data continuously, creating a constant stream of new information.

Smartphones and Mobile Devices
Smartphones are among the largest generators of personal data.
Depending on the permissions granted by the user, a smartphone may collect information such as:

- GPS location
- Photos and videos
- Contacts
- Text messages
- App usage
- Internet browsing activity
- Fitness and health information
- Calendar events
- Payment transactions

Many apps collect only the information needed to perform their functions, while others may request access to additional data. Understanding app permissions helps users make informed decisions about their privacy.

Online Activity
Every time you visit a website, information may be collected about your visit.
Examples include:

- Pages viewed
- Time spent on the website
- Search terms entered
- Products viewed
- Items placed in a shopping cart
- Purchases completed
- Device type
- Browser type

Organizations use this information to improve websites, recommend products, personalize content, and understand customer behavior.

Common Sources of Data Collection

SOURCE	EXAMPLES OF DATA COLLECTED
Smartphones	Location, app usage, photos, contacts
Online shopping	Purchases, payment methods, browsing history
Social media	Posts, likes, comments, shares
Streaming services	Viewing history, favorite genres
GPS navigation	Routes, traffic patterns, travel times
Wearable devices	Heart rate, sleep, exercise, step counts
Smart home devices	Temperature, energy use, voice commands

13.11: The Internet of Things (IoT)

The **Internet of Things (IoT)** refers to everyday devices that are connected to the Internet and can collect, share, and sometimes act on data.
Examples include Smart thermostats, video doorbells, smart watches, fitness trackers, connected appliances, smart lighting systems, agricultural sensors, industrial equipment, traffic sensors, and more.

These devices often communicate automatically without requiring direct human interaction.
For example, a smart thermostat may monitor indoor temperature, detect when people are home, and automatically adjust heating or cooling to improve comfort and reduce energy use.
As the number of connected devices grows, the amount of data they generate continues to increase dramatically.

Sensors

Many devices collect information using **sensors**.
A sensor detects physical conditions and converts them into digital data that computers can process.
Common sensors include Temperature sensors, motion sensors, light sensors, pressure sensors, proximity sensors, cameras, microphones, GPS receivers, fingerprint scanners, and more.

Sensors make it possible for computers to interact with the physical world.
For example:

- A fitness tracker measures heart rate.
- A weather station measures rainfall and wind speed.
- An automatic door opens when a motion sensor detects someone approaching.
- A smartphone rotates its screen using an internal motion sensor.

13.12: Data Quality

Why Data Quality Matters

Collecting large amounts of data is useful only if the data is accurate and reliable.
Data quality refers to how complete, accurate, consistent, and current data is. Poor quality data can result from:

- Typing errors
- Missing information
- Duplicate records
- Outdated information
- Faulty sensors
- Incorrect measurements

Poor data quality can lead to incorrect business decisions, medical mistakes, financial losses, or misleading research results.
For this reason, organizations spend significant time validating and maintaining their data.

Characteristics of High-Quality Data

CHARCTERISTICS	DESCRIPTION
Accurate	Correct and free from errors
Complete	Contains all necessary information
Consistent	Matches information across systems
Current	Up to date*
Reliable	Collected from trustworthy sources

*** Don't store data that will become out of date. Instead of age store the year born.**

Managing Data Responsibly

Collecting data is only the first step. Organizations must also manage data responsibly throughout its entire life cycle.
Good data management includes:

- Protecting sensitive information
- Backing up important data
- Controlling who has access
- Correcting errors
- Removing duplicate records
- Following privacy laws and organizational policies
- Disposing of data securely when it is no longer needed

Responsible data management helps organizations maintain trust while reducing the risk of security breaches and data loss.

Check Your App Permissions

Many mobile apps request access to features such as your location, camera, microphone, contacts, photos, or calendar.

Take a few minutes to review the permissions on your smartphone. Ask yourself:

- Does this app really need this information?
- Can I disable any permissions without affecting how the app works?
- Do I still use this app?

13.13: What Is Data Analytics?

Collecting large amounts of data is valuable only if the data can be used to answer questions or solve problems. **Data analytics** is the process of examining data to identify patterns, trends, and relationships that support better decision-making.
Organizations use data analytics to:

- Understand customer behavior.
- Improve products and services.
- Detect fraud.
- Predict future demand.
- Improve healthcare.
- Manage inventory.
- Optimize transportation.
- Support scientific research.

Instead of relying on guesses, organizations can use evidence from data to make more informed decisions. For example, a grocery store may analyze sales data to determine which products sell best before holidays. A hospital may analyze patient records to identify factors that improve recovery rates. A city may study traffic data to determine where additional traffic signals are needed.

The Four Types of Data Analytics
Although organizations use many different analytical techniques, most data analysis falls into four categories:

Descriptive analytics answers the question **What happened?**
Descriptive analytics helps organizations understand past performance.
Examples: Monthly sales reports, average exam scores, annual company revenue, number of website visitors.

Diagnostic analytics answers the question **Why did it happen?**
This type of analysis looks for relationships and possible causes.
Diagnostic analytics helps organizations identify factors that influenced results.
Examples: Why did sales decrease? Why did website traffic increase? Why did customer satisfaction improve?

Predictive analytics answers the question **What is likely to happen next?**
Using historical patterns and statistical models, predictive analytics estimates future outcomes.
Predictions are estimates rather than guarantees, but they help organizations prepare for future events.
Examples: Forecasting weather, predicting product demand

Prescriptive analytics answers the question **What should we do?**

This type of analytics recommends actions based on available information.
Examples include: Suggesting the fastest driving route, recommending inventory levels, identifying the best shipping method, helping doctors choose treatment options.
Prescriptive analytics often combines human expertise with computer-generated recommendations.

13.14: Data Visualization

Large collections of numbers can be difficult to understand.
Data visualization presents information using charts, graphs, maps, diagrams, and other visual formats that help people recognize patterns more quickly. Common visualizations include:

- Bar charts
- Line graphs
- Pie charts
- Maps
- Scatter plots
- Heat maps • Timelines

Well-designed visualizations allow decision-makers to understand complex information at a glance.

For example:

- A line graph can show changes in sales over time.
- A bar chart can compare sales across different products.
- A map can display weather conditions or delivery routes.
- A heat map can highlight areas with heavy traffic.

Good visualizations make data easier to interpret without changing the underlying information.

Technology in Practice

Read a Chart Carefully
Charts and graphs can make information easier to understand, but they can also be misleading if they are poorly designed or presented without context. When viewing a chart, ask yourself:

- What question is the chart answering?
- Where did the data come from?
- Is the scale appropriate?

- Are important details missing?
- Does the chart support the conclusions being presented?

Being able to interpret data visualizations is an important part of digital and information literacy.

13.15: Databases

Every day, organizations collect enormous amounts of information. Hospitals maintain patient records, banks store financial transactions, retailers manage inventory, streaming services remember viewing histories, and universities track student enrollment.
To keep this information organized and accessible, organizations use **databases** and **cloud storage systems**.
A well-designed storage system allows users to: Find information quickly, update records accurately, share information securely, support data analytics, protect valuable information.

As organizations continue collecting larger amounts of data, efficient storage becomes increasingly important.

Relational Databases

The most common type of database is a **relational database**.
A relational database organizes information into **tables** made up of rows and columns.

- **Rows** represent records.
- **Columns** represent fields.

Related tables can be connected using common values, allowing information to be shared without unnecessary duplication.

Relational Database for Small Company

The example below shows a database that has 2 tables: **Employee** and **Department**.

Employee Table:

The **Employee table** has Employee ID, Last Name, First Name, Full time (Y or N), hourly rate, and department code.

Employee ID	Last Name	First Name	Full time	hourly rate	Department
E3258	Jones	Carl	Y	15.95	A1
E3258	Smith	Mary	Y	10.95	A1

You can see from this that the company has 2 full time employees: Carl Jones and Mary Smith. They both have a value of "A1" for department.

Department Table:

The **Department table** has Dept ID, Title and Director:

Dept ID	Title	Director
A1	Accounting	E3258

You can see from this table that the company has just one department.

Now, if you look at the two tables together you can see that the two tables are related: Mary Smith is in the Department A1 which is Accounting. Her boss is "E3258" which is Carl Jones.

If we want to print a list of all employees, we can use the department from the **Employee** table to get the **Title** of the department from the **Department** table.

If we want to print a list of everyone in accounting department, we can look in the Employee table for a department of "A1"

Instead of storing information repeatedly in every table, the database uses unique **Employee ID** and **Dept ID** to connect related information.

More important is that if the director of the Accounting department changes, we only have to change it in one place. This organization improves efficiency and reduces errors.

Together, these related tables form a complete company database.

This is a very small company, thinking big! A large organization might have a much larger database: tables for products, suppliers, customers, contractors, etc.

13.16: Cloud Databases

Traditionally, organizations stored databases on computers located within their own buildings. Today, many organizations use **cloud databases**, which store information on servers operated by cloud service providers and accessed through the Internet.

Cloud databases allow organizations to:

- Access information from multiple locations.
- Support remote employees.
- Expand storage as needed.
- Improve collaboration.
- Reduce the need to maintain their own servers.

Examples: Online banking systems, streaming services, online retailers, learning management systems such as Blackboard, social media platforms

13.17: Data Lakes: Storage for Unstructured Data

A **data lake** is a large storage system designed to hold enormous amounts of data in its original form. Unlike a traditional database, which usually organizes information into structured tables, a data lake can store **unstructured and semi-structured data** such as photographs, videos, audio recordings, social media posts, documents, sensor readings, and log files. Organizations can collect this data first and decide later how to organize and analyze it. Data lakes are especially useful in **big data, artificial intelligence, and machine learning**, where large amounts of many different types of data may be needed.

13.18: Free Research Databases

The table below shows some free searchable databases used by researchers:

Free Database	Organization	What Data is Available	Example Question
Bureau of Labor Statistics (BLS)	U.S. Department of Labor	Employment, unemployment, wages, occupations, inflation, productivity	*What is the average salary for a registered nurse?*

data.census.gov	U.S. Census Bureau	Population, income, education, housing, employment,	*What percentage of people in my*
Free Database	**Organization**	**What Data is Available**	**Example Question**
		demographics	*county have a bachelor's degree?*
Data.gov	U.S. Government	Hundreds of thousands of government datasets covering health, climate, education, transportation, agriculture, and more	*How has electric vehicle use changed in the United States?*
PubMed	National Library of Medicine	Medicine, health, biology, public health, and biomedical research	*What does research say about exercise and heart health?*
ERIC	U.S. Department of Education	Education research, teaching methods, learning, colleges, schools, and educational technology	*What does research say about online learning?*
World Bank Open Data	World Bank	Population, poverty, education, economics, health, environment, and international development	*How does life expectancy differ among countries?*
CDC Data and Statistics	Centers for Disease Control and Prevention	Public health, disease, injuries, mortality, nutrition, and health behaviors	*How have smoking rates changed over time?*
National Center for Education Statistics (NCES)	U.S. Department of Education	Schools, colleges, enrollment, graduation, tuition, degrees, and educational outcomes	*How has college enrollment changed over the past 20 years?*