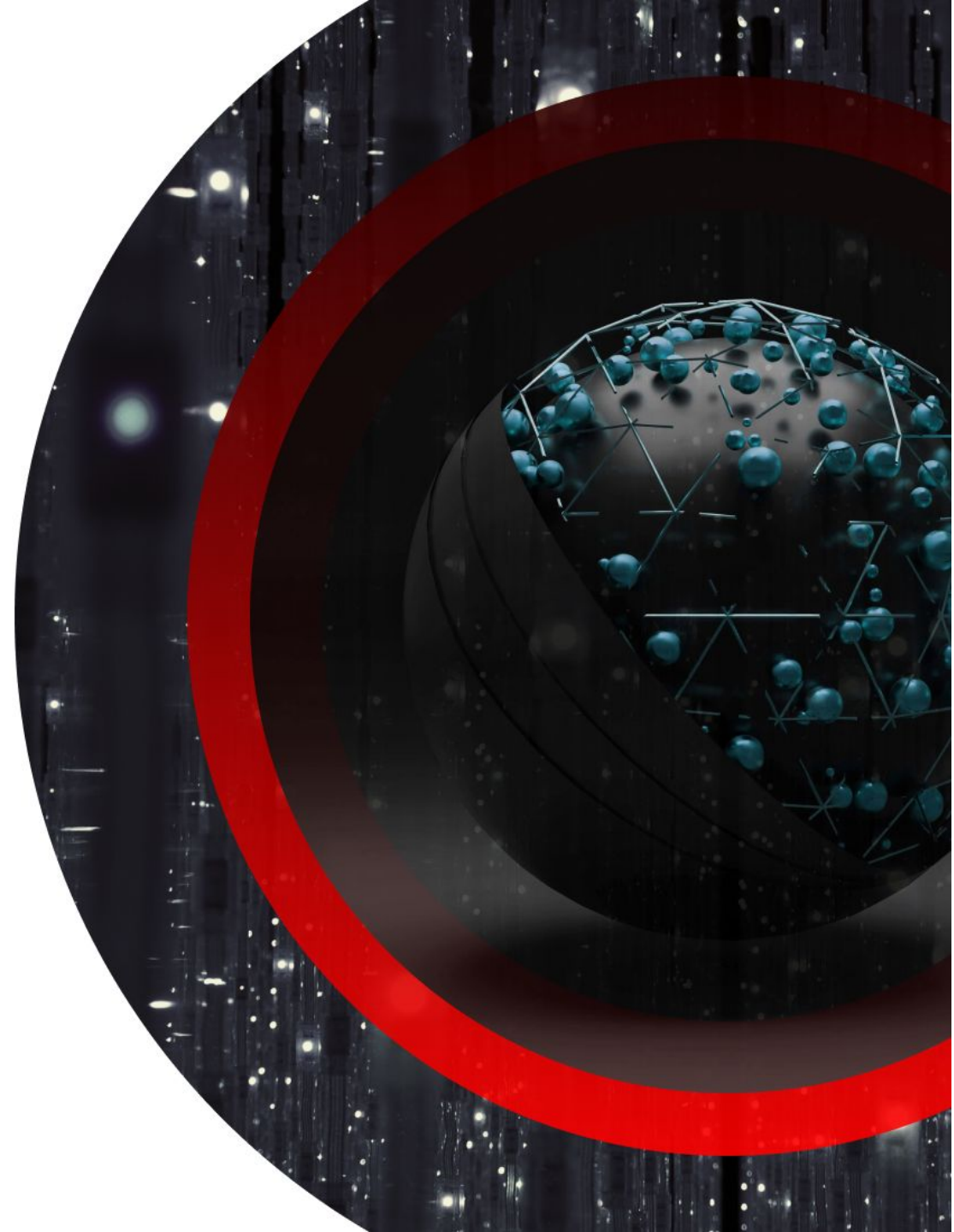


DATA SCIENCE AND ANALYTICS

Introductory Course



CLASS NORMS

COURTESY IN CLASS

Remaining on mute unless called on, exercising courtesy during breakout rooms, using the chat box for questions only

ATTENDANCE

100% attendance is expected and contributes to success in passing the course and the program.

PARTICIPATION

Keeping an open mind in discussions and sharing experiences, making contributions during team assignments, submitting assignments in Canvas, and participating in discussion boards

USE OF CLASS RESOURCES

Follow along during the lecture with the lesson companion and download any in-class documents prior to class.





PROGRAM PATH

1 Introductory Course

2 SQL and Databases

3 Statistics and Probability

4 Data Storytelling

Milestone 1: Building and
Presenting Data Stories

5 Python Programming

6 Data Wrangling

7 Visual Communications

8 Advanced SQL Programming

Milestone 2: Data Integration,
Preparation, Reporting, and
Presentation

9 Business Intelligence

10 Big Data

11 Machine Learning

12 Applied AI

Milestone 3: Capstone Project –
Delivering Insights and
Presentations

INTRODUCTORY COURSE PATH

1

INTRODUCTION TO
DATA SCIENCE AND
ANALYTICS

2

**COMPUTING
PRIMER**

3

PROGRAMMING
CONCEPTS

4

DISCOVERING AND
CURATING DATA

5

STRUCTURING AND
ANALYZING DATA

6

CLEANING AND
ENRICHING DATA

7

VALIDATING AND
PRESENTING DATA

8

INTRODUCTION TO
DATA SCIENCE PROJECTS

9

ASSESSMENT NIGHT





LESSON 1: COMPUTER MEMORY.



LESSON OUTLINE.

- Identify the basic parts of a computer, including RAM, the hard drive, and the CPU.
- Understand different units of memory.
- Describe the differences between a computer's RAM and hard drive.





DISCUSSION

Think of a time when you were shopping for a new laptop/computer. What factors did you consider?





LESSON GOALS



WHAT ARE THE GOALS?

Identify the basic parts of a computer, including RAM, the hard drive, and the CPU.

Understand how computer memory is measured.

Describe the differences between a computer's RAM and hard drive.



WHY ARE THEY IMPORTANT?

Computers are the primary tools used by data professionals in their work.

It's important that data professionals have a solid understanding of how computers process data and perform calculations.



YOUR TAKE

What are your goals for the lesson?

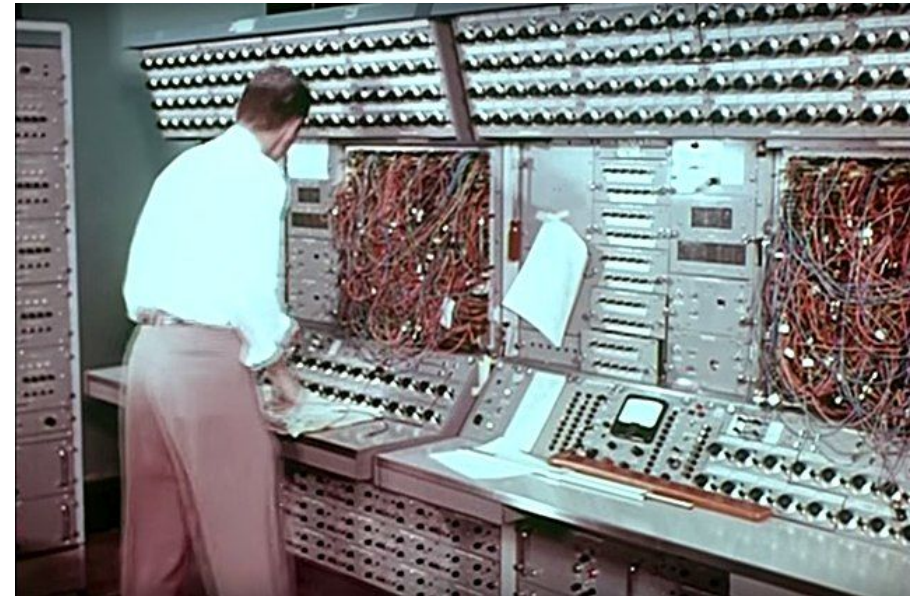
Share three things you know about the lesson topic and two things you want to know about the lesson topic.





WHAT IS COMPUTING?

The use of a computer to process data or perform calculations





SOFTWARE VS. HARDWARE

Data can be stored in two different places in a computer: programs and physical components

Programs on a computer are referred to as **software**, while the physical components of a computer are referred to as **hardware**.





COMPUTER HARDWARE

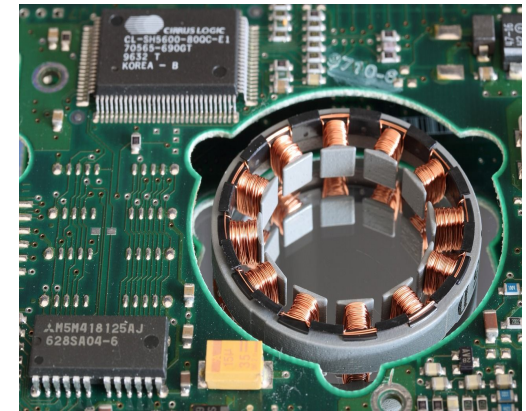
- Central Processing Unit (CPU)
- Random Access Memory (RAM)
- Hard Drive





CENTRAL PROCESSING UNIT

- The CPU, also known as the processor, provides the computer with power and instructions for it to complete tasks.
- Without a processor, a computer would not be able to function.
- The CPU does not store any information.



[Pixabay](#)



RANDOM ACCESS MEMORY.

- RAM is short-term memory that stores information that is currently being used by a computer.
- RAM allows a computer to switch quickly between programs being run.
- The more RAM a computer has, the more tasks it can manage at one time.





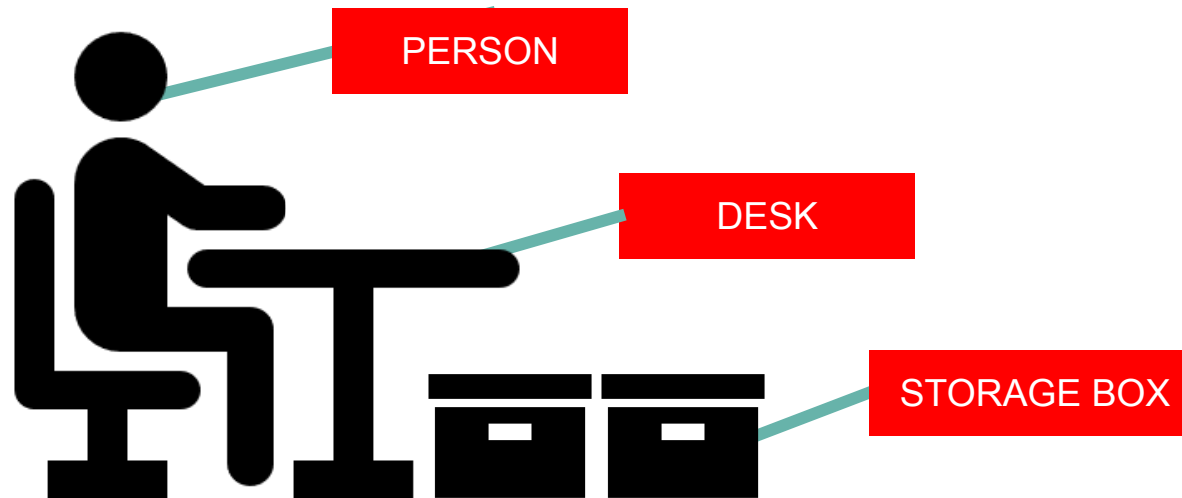
HARD DRIVE.

- A device used to store data on a computer
- The hard drive is long-term storage. It contains everything from personal files to the information needed for operating systems to run.



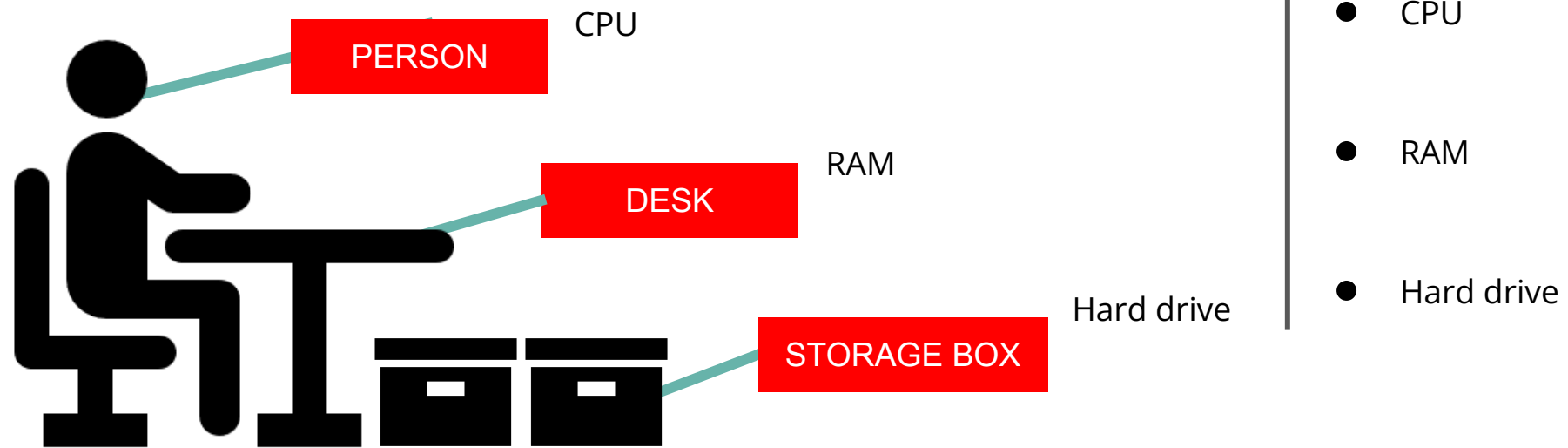


ACTIVITY: MATCH FUNCTIONS



- CPU
- RAM
- Hard drive

ACTIVITY: FEEDBACK





BINARY CODE.

- Binary is the language a computer reads.
- A series of 0s and 1s
- Computers can only read 0s and 1s.

01010111	01101001	01101011
01101001	01110000	01100101
01100100	01101001	01100001



BINARY: TWO-SYMBOL SYSTEM



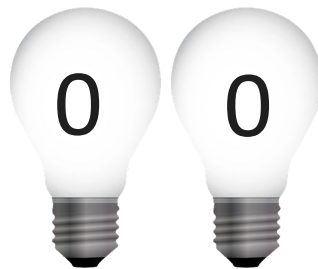
"0" represents no flow of electricity.



"1" represents electricity being allowed to flow.



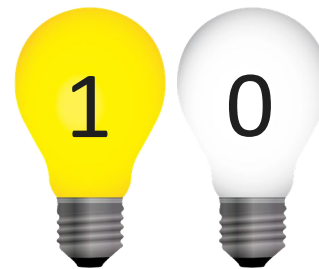
LIGHT BULBS SCENARIOS



Scenario 1



Scenario 2



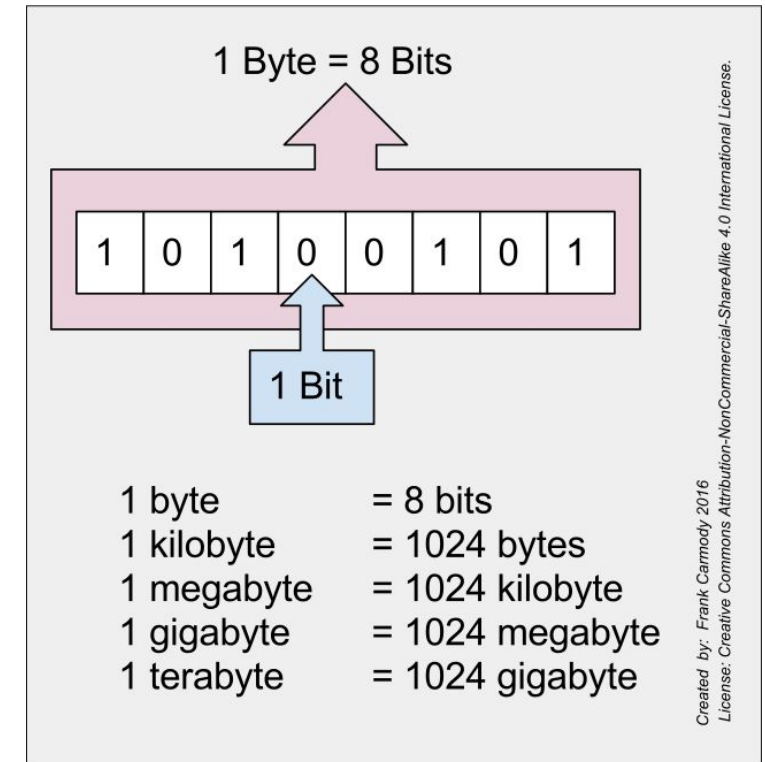
Scenario 3



Scenario 4

BINARY DIGITS: BITS AND BYTES.

- At the smallest scale in the computer, information is stored as **bits** and **bytes**.
- Memory is measured in bits and bytes.
- B = byte, b = bit
- 1 byte = 8 bits





MEASUREMENTS

kb = kilobit

kB = kilobyte

mb = megabit

mB = megabyte

gb = gigabit

gB = gigabyte

tb = terabit

tB = terabyte

1 kilobyte = 1,024 bytes

1 megabyte = 1,024 kilobytes

1 gigabyte = 1,024 megabytes

1 terabyte = 1,024 gigabytes



Why 1024?

Normally you use the prefix of *kilo* to represent 1000 of something. However, the computer field adopted the convention of using powers of 2 rather than powers of 10 to measure units. When we talk about computer units, a kilobyte is 2 to the power 10 bytes.



POLL



POLL:

Take a guess: Approximately how many songs can be stored using 1 GB of memory?

- A. 75
- B. 250
- C. 500





POLL: FEEDBACK.

Approximately how many songs can be stored using 1 GB of memory?

- A. 75
- B. 250**
- C. 500

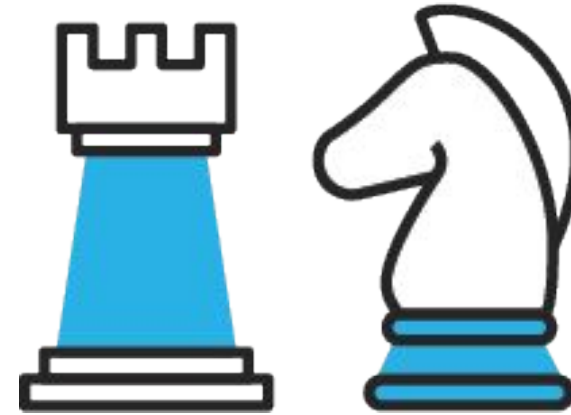




1.2.1 ACTIVITY.

Maisie has decided to start a business selling mugs with cool internet slang and memes. She is currently using a 10-year-old computer and needs a new computer for her online store.

Answer the questions in the 1.2.1 Activity Google Doc to determine the best computer for Maisie's new business.





REVIEW AND WRAP-UP

In this lesson we learned to:

- Identify the basic parts of a computer, including RAM, the hard drive, and the CPU.
- Understand different units of memory.
- Describe the differences between a computer's RAM and hard drive.





YOUR TAKE

What have you learned so far?

Share one thing you learned about the lesson topic.



QUESTIONS?





**BREAK
TIME.**



LESSON 2: THE INTERNET, CLOUD COMPUTING, AND NETWORKS.



LESSON OUTLINE.

- Define what a network is and understand how it works on a basic level.
- Identify different types of networks and how they differ from each other.
- Explain the advantages and disadvantages of cloud computing.
- Identify known security threats and common measures used to combat them.





DISCUSSION

What steps do you take to keep your sensitive information safe while online?





LESSON GOALS



WHAT ARE THE GOALS?

Define what a network is and understand how it works on a basic level.

Identify different types of networks and how they differ from each other.

Explain the advantages and disadvantages of cloud computing.

Identify known security threats and common measures used to combat them.



WHY ARE THEY IMPORTANT?

Every company uses networks in some capacity. Especially as companies begin to transition to cloud computing, it's important to have a solid understanding of how the internet works and its vulnerabilities.



YOUR TAKE

What are your goals for the lesson?

Share three things you know about the lesson topic and two things you want to know about the lesson topic.





WHAT IS A COMPUTER NETWORK?

A computer network is a group of computers that are connected.

- Networks range in size from just two computers to millions of computers.
- Networks are connected using either physical cables or wireless connections.
- Smaller networks communicate with other networks using a router.



TYPES OF NETWORKS.

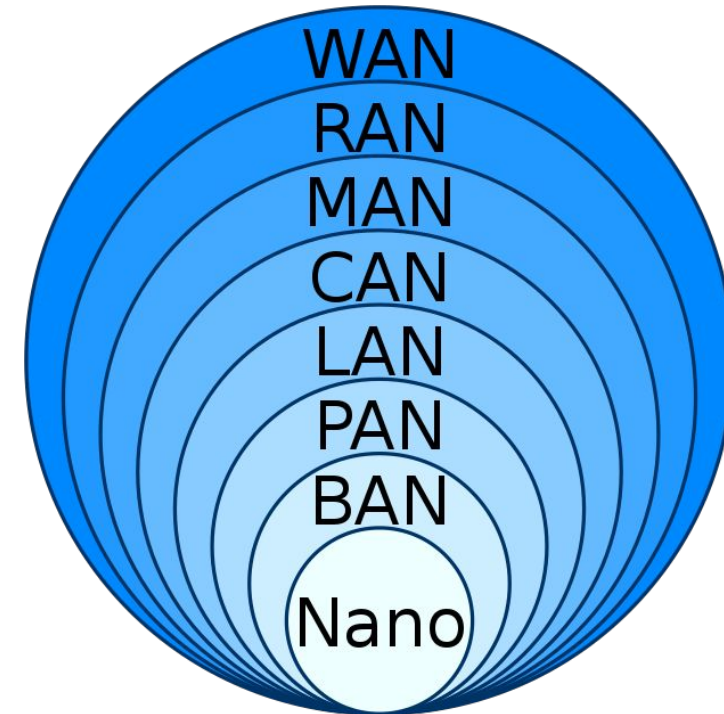
Geographic location often defines a computer network.

Common categories of networks include:

PAN: Personal Area Network

LAN: Local Area Network

VPN: Virtual Private Network





KNOWLEDGE CHECK!

Scenario

You were asked to build a network that allows employees to share access to a printer and a group messaging system.

What type of network would you choose?

- A. PAN (Personal Area Network)
- B. LAN (Local Area Network)
- C. VPN (Virtual Private Network)



KNOWLEDGE CHECK: FEEDBACK.

Scenario

You were asked to build a network that allows employees to share access to a printer and a group messaging system.

What type of network would you choose?

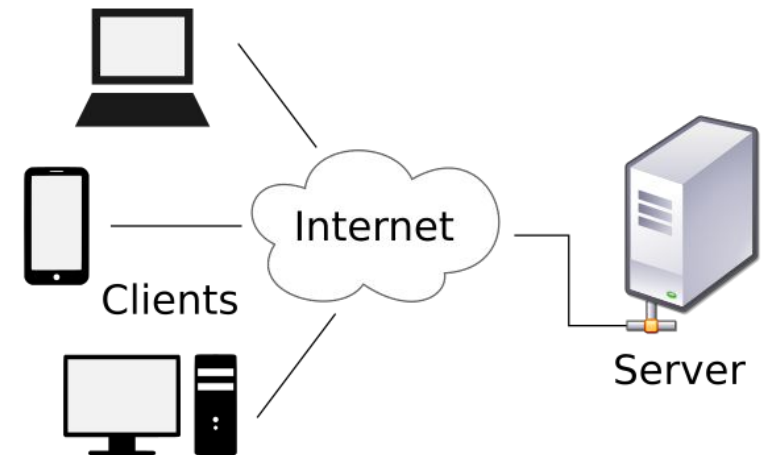
- A. PAN (Personal Area Network)
- B. LAN (Local Area Network)**
- C. VPN (Virtual Private Network)

NETWORK ARCHITECTURE

Network architecture falls into one of two categories.

P2P (Peer to Peer): All devices connected on a network have equal privileges on the network.

Client-Server: A client (device) requests a resource from a server that responds and provides the requested resource.





IP ADDRESS

A unique number given to a device connected to a network is known as an IP address.

IPv4: 100.200.100.200

IPv6: 2002:64C8:64C8::



INTRANET AND INTERNET

Intranets are private networks that connect a group of computers.

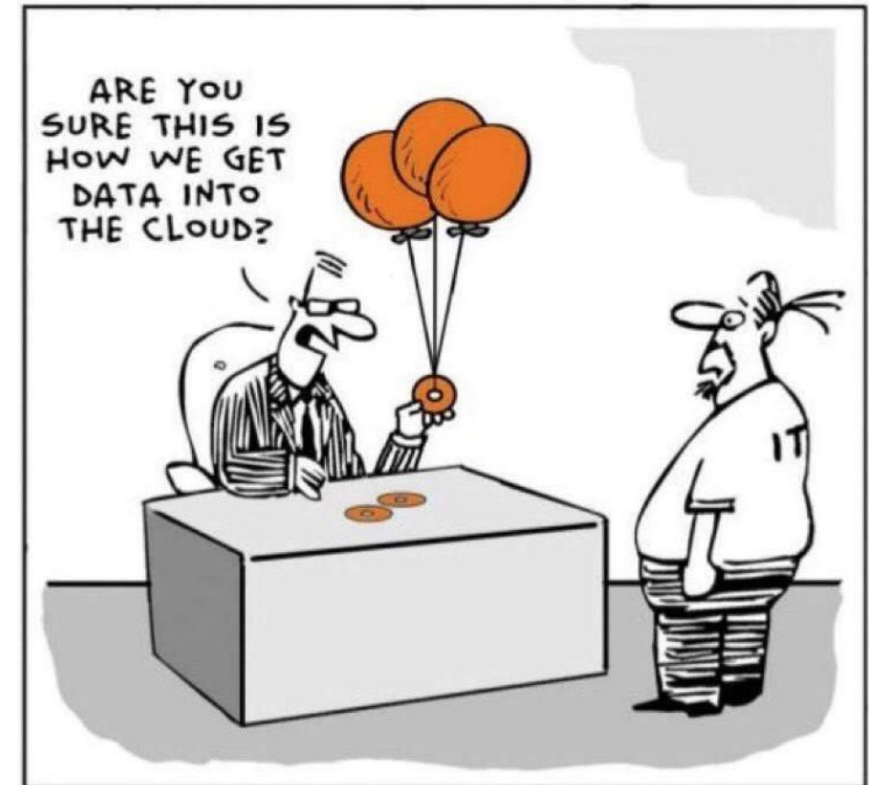
- Intranets are typically used by companies to enable internal communication and collaboration.
- Only authorized devices can connect to the network.
- Intranets are suitable for large companies with many employees.

The internet is the world's largest computer network, commonly referred to as a "network of networks."

- It functions like any other network and uses the client-server model.

WHAT IS CLOUD COMPUTING?

The cloud is a framework in which data is stored on servers over the internet instead of on a hard drive or internal server.



From: [Alteryx](#)





THE CLOUD.

Advantages

- Access from any device with internet access
- No risk of losing data if hard drive fails
- Multiple people can access the same data source simultaneously.
- Less expensive than investing in your own server

Disadvantages

- No internet connection = no access to data
- Service outages by the cloud provider
- Third party is responsible for your data.
- Cyberattacks



NETWORK SECURITY.

Networks can be infiltrated by bad actors who are looking to steal information, get money, or just cause chaos.

There must be a balance between making data accessible and protecting it from outside actors.





CYBERATTACKS.

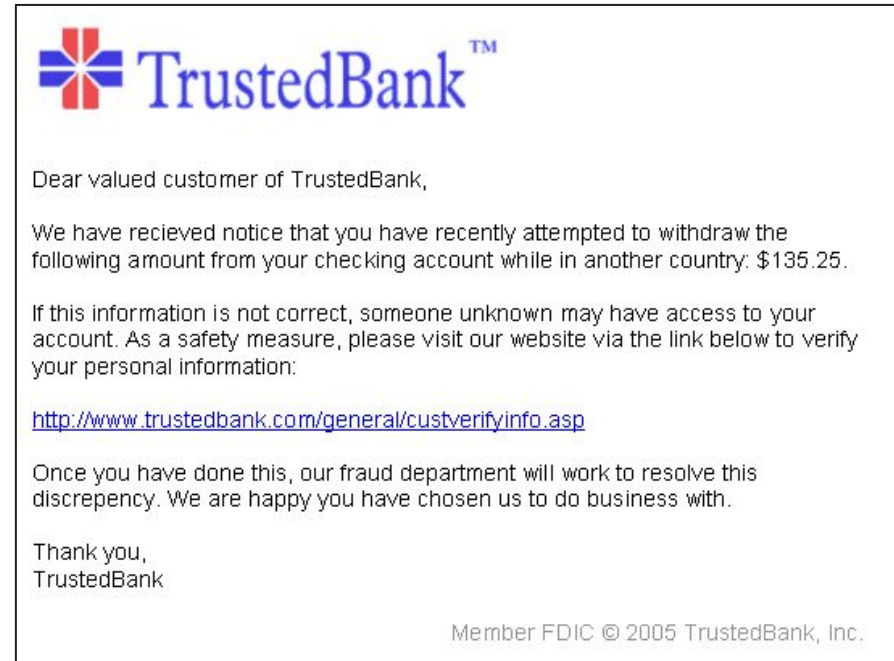
There are many ways that bad actors can infiltrate or take down a network:

- Trojan horse
- DoS
- Ransomware
- Password attack
- Phishing



PHISHING.

Fraudulent link



From: [Wikipedia](#)





INTERNET USER PRIVACY.

User information revealed by the IP address:

- Internet service provider, approximate location, time zone, and even the browser and operating system

Ways to mask user information:

- VPN
- Proxy server

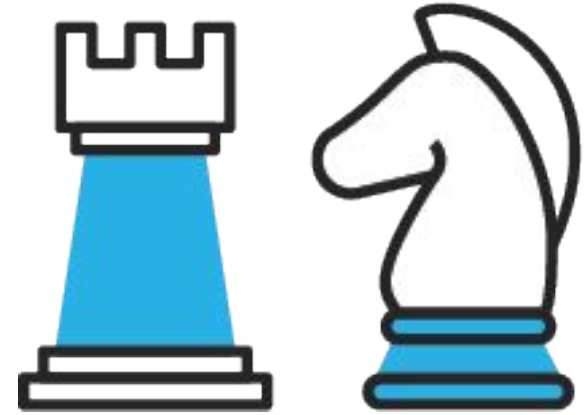




1.2.2 ACTIVITY.

While it's true that the internet can be accessed from anywhere in the world, not everyone experiences the internet the same way.

Follow the steps in the 1.2.2 Activity Google Doc to learn more about how the world experiences the internet in different ways.





DISCUSSION

Does this lesson change your feelings about transmitting your sensitive information online?





REVIEW AND WRAP-UP

In this lesson we learned to:

- Define what a network is and understand how it works on a basic level.
- Identify different types of networks and how they differ from each other.
- Explain the advantages and disadvantages of cloud computing.
- Identify known security threats and common measures used to combat them.





YOUR TAKE

What have you learned so far?

Share one thing you learned about the lesson topic.



QUESTIONS?





**BREAK
TIME.**

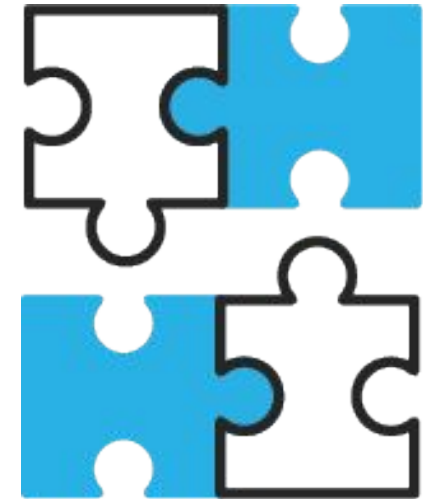


LESSON 3: SCALING AND BIG DATA



LESSON OUTLINE:

- Identify scaling issues within the physical and virtual infrastructure.
- Describe techniques used to scale infrastructure.
- Discuss how big data has changed the way businesses operate.





DISCUSSION.

Have you ever found yourself in a situation in which your computer was not powerful enough to do what you wanted? How did you overcome this challenge?





LESSON GOALS



WHAT ARE THE GOALS?

- Identify scaling issues within the physical and virtual infrastructure.
- Describe techniques used to scale infrastructure.
- Discuss how big data has changed the way businesses operate.

WHY ARE THEY IMPORTANT?

- Businesses are increasingly working with very large sets of data. More data means more computing power is needed.
- It is important for data professionals to understand the challenges of scaling data infrastructure and common scaling methods used to solve these problems.



YOUR TAKE

What are your goals for the lesson?

Share three things you know about the lesson topic and two things you want to know about the lesson topic.





BIG DATA.

Big data is large, complex amounts of data. It can be organized or disorganized.





SCALING

The ability for systems to perform as workload increases is called **scalability**.

- Businesses are always growing, so technology must be able to grow with it.
- Technology may eventually become outdated or no longer powerful enough.
- Data professionals need to understand which strategies work best in each situation.





IMPORTANCE OF SCALING

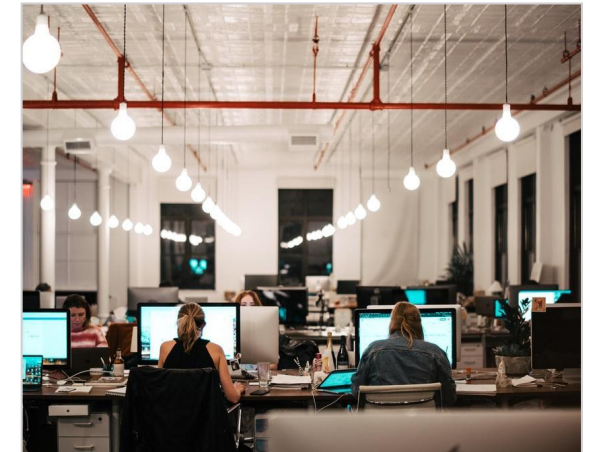
Scaling is when additional parts and tools can be added to the data infrastructure without starting over.

- Contributes to the efficiency of a company
- Saves time and money
- Minimizes repetitive work



WHEN TO SCALE

- **Hardware:** The physical components of the data infrastructure are unable to support the increased workload.
- **Memory:** Systems not having enough memory to complete the task
- **Network Congestion:** Reductions or failures in service due to a system trying to process too much data



From: [Israel Andrade on Unsplash](#)



HOW TO SCALE

Vertical scaling

Improving performance by using more advanced software and/or hardware within existing systems

Horizontal scaling

Improving performance by adding or using more software and/or hardware



VERTICAL SCALING: PROS AND CONS

Advantages

- Easier to implement
- More cost efficient than horizontal scaling
- Requires fewer resources

Disadvantages

- Infrastructure failures can have larger consequences.
- Horizontal scaling may eventually be required anyway.



HORIZONTAL SCALING: PROS AND CONS

Advantages

- Long-term solution
- More systems within the infrastructure mean failures in one system have fewer consequences.

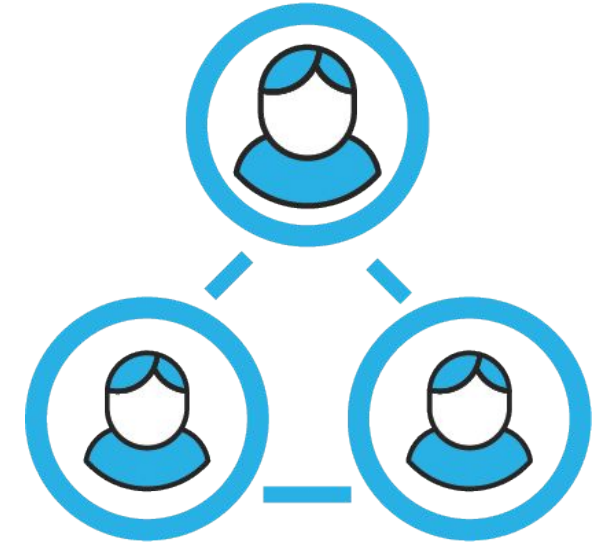
Disadvantages

- Usually less cost-effective
- Requires more resources



BRAINSTORMING

Describe a scenario in which horizontal scaling may be preferable to vertical scaling or vice versa.





1.2.3: ACTIVITY.

Help Big AI decide how to scale up his business and handle an increased workload. You'll use the 1.2.3 Activity Google Doc and the **1.2.3 Activity Dataset.xlsx** file to complete this activity.



[Pixabay](#)



INTERVIEW TIME!

What is the difference between vertical and horizontal scaling?





REVIEW AND WRAP-UP

In this lesson we learned to:

- Identify scaling issues within the physical and virtual infrastructure.
- Describe techniques used to scale infrastructure.
- Discuss how big data has changed the way businesses operate.





YOUR TAKE

What have you learned so far?

Share one thing you learned about the lesson topic.



QUESTIONS?





NEXT STEPS



Assigned Activities



Reminders

